

Empirical Foundations
of
Educational
Research

Empirical Foundations of Educational Research

Gilbert Sax

*Professor of Educational Psychology
University of Washington*

PRENTICE-HALL, INC., *Englewood Cliffs, New Jersey*

© 1968 by Prentice-Hall, Inc., Englewood Cliffs, New Jersey. All rights reserved. No part of this book may be reproduced in any form or by any means without permission in writing from the publisher.

Library of Congress Catalog Card Number: 68-10609

Printed in the United States of America

Current Printing (last digit):

10 9 8 7 6 5 4 3

Prentice-Hall International, Inc., *London*
Prentice-Hall of Australia, Pty. Ltd., *Sydney*
Prentice-Hall of Canada, Ltd., *Toronto*
Prentice-Hall of India Private Ltd., *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*

To my parents
and children
Laurie, Kathy, and Karen,
but especially
to Judy

Preface

The application of empirical research methods to education goes back little more than half a century. Since then, the schools have been called upon to fulfill roles and obligations largely unconceived of by nineteenth century educators. Unfortunately, it has not always been possible to justify modern educational practices or innovations. Like his predecessor, today's educator is often forced into making decisions without knowledge derived from empirical research findings to guide him or to help justify his actions.

Current knowledge about education has come from many sources, ranging in validity from the biased pronouncements of the grossly misinformed to highly dependable statements originating with those who have gone to the trouble to test their expectations empirically. Research methods are devised to provide investigators with the techniques, procedures, and knowledge needed to verify or reject these expectations, not with the hope of "proving" one's own point or in finding fault with others', but to study problems within a scientific framework that encourages the resolution of problems by rational and empirical means.

This text is designed to provide readers with the empirical foundations of educational research. It is deliberately not a textbook in statistics, measurements, or historical methodology, although all of these areas are, of course, related to empirical research methods.

I am convinced that to try to explain research design to students who have not had at least an introductory course in statistics is to do them a great disservice. They are likely, in a one-semester course in research methods, to learn neither statistics nor research design and to dislike both. Unless graduate students in education are given training in statistics and research design, we are contributing to the unhappy precedence of preparing students who are unable to read professional journals or evaluate their own practices. Education will not be regarded as a profession until its practitioners have mastered these foundational skills.

Nor does this text pretend to cover historical or philosophical research

methods. Historiography is in itself a highly complex field deserving to be studied on its own by those planning to specialize in the history or philosophy of education. The few pages on historical research methods typically included in texts on research does justice neither to history nor to empiricism.

This text has the following purposes:

1. To acquaint graduate students in education with the potentials and limitations of empirical research;
2. To provide students with an understanding of the contributions that empirical research methods have made to educational practices;
3. To show the relationships between empiricism and the philosophical and theoretical assumptions underlying empirical research procedures;
4. To describe the tools and methods used in empirical investigations.

To accomplish these objectives, this text has incorporated a number of special features:

1. Chapters are arranged sequentially, taking the student from the point of selecting a research project to the analysis and presentation of his findings;
2. Considerations related to the writing of a research report are included throughout the text and not in a separate section or appendix. The research report is a formal way of communicating scientific information, and as such, its contents should reflect these understandings. By keeping related elements together, these relationships can be most easily seen;
3. A rather detailed summary is provided at the end of each chapter. This can serve two purposes: to introduce the student to the contents of the chapter and to help him to review its contents;
4. Every chapter contains an annotated bibliography. It is provided to take the curious reader into greater depth than can be presented in this text;
5. Practice exercises accompany each chapter. Some exercises can easily be converted to objectively scored examinations; others are designed as projects or for class discussion. Difficulty levels range widely. The majority of the exercises are designed to test how well the student can apply and understand the concepts he has read about;
6. The text is specifically designed for master's degree candidates in education, although it may also be used with more advanced students if the supplementary readings are used extensively;
7. The assumption is made that the student has taken and understands

basic elements of descriptive and sampling statistics. If this assumption cannot be met, it is suggested that the instructor cover these topics as early in the semester as possible:

8. The text is suitable for either a one-semester or year course in research methods. For a one-semester course, the student can read approximately one chapter per week and perhaps design a study of his own; for a more extended course, each chapter can be enriched with supplementary readings. Students might also be asked to actually complete their theses if the course can be carried out over a year's time.

I am indebted to a number of persons who were extremely helpful in the preparation of this manuscript: Mrs. Jane Fujitani typed all corrections in the original and in subsequent copies; Miss Gayle Ouye typed the final manuscript; Mrs. Virginia Kemble, Miss Ethel Oda and Mrs. Nona Springel were responsible for checking bibliographical references for accuracy and completeness; Miss Cathy Wentworth read the final page proofs; and I am indebted to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., Cambridge, and to Oliver & Boyd Ltd., Edinburgh, for their permission to reprint from their book *The Design of Experiments*. To all these persons, I owe a debt of gratitude.

Various parts of the text were read and critiqued by Dr. David G. Ryans, Dr. George Fujita, Dr. Robert Clopton, Dr. Donald Fitzgerald, Dr. Donald Leton, Dr. Julian C. Stanley, and Dr. Ted Husek. Dr. Agnes Niyekawa read the entire text and made numerous helpful suggestions. To all of these persons, I am deeply grateful. They are, of course, not responsible for any errors of omission or commission which may have occurred.

Writing a textbook requires that an author neglect his family and friends. To them all I owe a special debt of gratitude and it is to them that I dedicate this text.

GILBERT SAX

Seattle, Washington

Contents

ONE: The Scientific Basis of Education 1

Science as a source of knowledge, 6; Characteristics of science, 7; Some assumptions made in science, 17; Some possible objections to a science of education, 20; Descriptive and explanatory levels of science, 22
SUMMARY, 25; PRACTICE EXERCISES, 28; SELECTED RELATED READINGS, 30

TWO: The Role of Research in Education 32

Types of research activities and their relationship to educational change, 33; Types of research and their relationship to innovation, 35; The historical development of educational research, 37; Factors hindering educational research, 52
SUMMARY, 60; PRACTICE EXERCISES, 63; SUPPLEMENTARY READINGS, 65

THREE: Selecting a Research Problem 67

Research to clarify and validate basic theoretical problems, 71; Research to correct methodological errors, 72; Research to resolve practical field problems, 75; Published suggestions for selecting research topics, 77; Delimiting and specifying the problem, 80
SUMMARY, 83; PRACTICE EXERCISES, 84; SELECTED RELATED READINGS, 85

FOUR: Reviewing the Literature 87

Writing the review of the literature, 88; Developing note-taking skills, 92
SUMMARY, 96; PRACTICE EXERCISES, 97; SELECTED RELATED READINGS, 98

CONTENTS

FIVE: The Research Hypothesis 100

Reasons for constructing hypotheses, 102; Criteria for evaluating hypotheses, 104; Some objections to hypothesis construction, 108; The logical form of the research hypothesis, 110; Writing the research hypothesis, 112; Definition of terms, 116

SUMMARY, 121; PRACTICE EXERCISES, 124; SUPPLEMENTARY RELATED READINGS, 126

SIX: Sampling Theory and Procedures 128

Determining the characteristics of the sample, 129; Methods of selecting samples, 130; Determining the size of the sample, 142

SUMMARY, 149; PRACTICE EXERCISES, 150; SUPPLEMENTARY READINGS, 152

SEVEN: Reliability and Validity of Psychometric Measures 154

The meaning of reliability, 157; Types of reliability coefficients, 158; Factors affecting reliability coefficients, 162; Required magnitudes of reliability coefficients, 165; Validity of measurements, 165; Types of validity coefficients, 166; Some difficulties in the validation of measurements, 170; Criterion contamination, 171; Cross validation, 171

SUMMARY, 172; PRACTICE EXERCISES, 173; SELECTED RELATED READINGS, 175

EIGHT: Some Problems of Observation 177

Sources of observational distortion, 177; Some general methods of reducing observational errors, 180; Critical incident technique, 191

SUMMARY, 196; PRACTICE EXERCISES, 198; SUPPLEMENTARY RELATED READINGS, 199

NINE: Obtaining Information from Respondents: I—the Interview and the Questionnaire 201

The interview, 201; The questionnaire, 214

SUMMARY, 233; PRACTICE EXERCISES, 237; SELECTED RELATED READINGS, 239

TEN: Obtaining Information from Respondents:
II—Other Selected Techniques 241

Disguised-nonstructured methods, 241; Disguised-structured methods, 258; Sociometry, 263; The Q-sort technique, 267; The semantic differential, 269; Content analysis, 273

SUMMARY, 279; PRACTICE EXERCISES, 238; SUPPLEMENTARY RELATED READINGS, 286

ELEVEN: Descriptive Research 288

The case study, 288; The sample survey, 292; Correlational studies, 292; Developmental studies, 309; Cross-cultural studies, 316

SUMMARY, 324; EXERCISES, 328; SELECTED RELATED READINGS, 331

TWELVE: Experimental Design 335

Some arguments against experimental methods in educational research, 337; Some criteria for evaluating experimental designs, 340; Experimental control, 346; Types of experimental designs, 360; Writing the chapter on experimental procedures, 377

SUMMARY, 378; PRACTICE EXERCISES, 382; RELATED SUPPLEMENTARY READINGS, 387

THIRTEEN: Data Processing 388

Data processing forms and records, 389; Computational aids, 400; The use of electronic computers, 402; Types of electronic computers, 405; Preparing data for electronic processing, 410; Writing the section on data analysis, 418

SUMMARY, 424; PRACTICE EXERCISES, 426; RELATED SUPPLEMENTARY READINGS, 428

Empirical Foundations
of
Educational
Research

The Scientific Basis of Education

At no time in man's history has the need for more reliable knowledge concerning education been as keenly felt as it is today. The schools, at one time agencies dedicated to the education of the elite, have been changing their functions, their curricula, and their methods to accommodate ever-expanding demands from a society that has insisted upon quality education for all of its children.

It is not surprising, considering the many changes in education over the past half century, that educators have lacked the knowledge needed to justify these modifications in educational practice. Instead, many justifications were based, at least in part, on the biases of individual educational reformers, on appeals to authority, or, perhaps even worse, on dogmatic assertion. Under such conditions it was easy for subjectivity and bias to override objective and empirical methods of studying educational problems.

This lack of valid knowledge and the concomitant dependence on prescientific ways of thinking not only allowed unexamined practices to continue but also prevented a number of worthwhile educational experiments from being conducted. In 1897, for example, J. M. Rice asked the participants at the annual meeting of the Department of Superintendence how it would be possible to determine if students who are given 10 minutes of spelling each day learn any less than those who are given 40 minutes each day. He discovered that some people in education did not even believe an answer to this question was possible. Rice stated, "... to my great surprise, the question threw consternation into the camp. The first to respond was a very popular professor of psychology engaged in training teachers in the West. He said, in effect, that the question was one which could never be answered; and he gave me a rather severe drubbing for taking up the time of such an important body of educators in asking them silly questions."¹

¹ J. M. Rice, *Scientific Management in Education* (New York: Noble & Noble, Publishers, Inc. 1914) pp. 17-18.

Dependable knowledge about education has come from many different sources. At the beginning of the twentieth century educators could not easily imagine that scientific methods could provide answers to questions that had interested man for the preceding two thousand years, and it was little wonder that Rice had difficulty persuading the members of the Department of Superintendence to accept these methods. The prevailing attitude was that educational problems could be resolved through appeals to five sources: common sense, authority, intuition, revelation, or "reason."

The Appeal to Common Sense

The appeal to common sense can be considered from two points of view: as a means for "justifying" preconceived beliefs or as a way of referring to empirically verified knowledge that has been generally accepted. When used to justify preconceived ideas, the appeal to common sense stifles communication and prevents or discourages experimentation. Too often, scientific investigation is considered unnecessary as long as common sense provides justification for our beliefs. Thus, as long as common sense dictated that the study of difficult subjects such as Latin, Greek, or mathematics would "exercise the mind," there were few demands for experimental investigation of the matter. Indeed, in 1892, when President Eliot of Harvard was appointed chairman of the Committee of Ten on Secondary School Studies by the National Educational Association, the Committee recommended that all subjects in the curriculum be accepted as equivalent in value. The Committee further stated that only difficult subjects could be counted on to "exercise the mind." The implementation of the Committee's recommendations led to a sterile, difficult, and impractical curriculum which did much to eliminate many of the less able students from the public schools. These "common sense" beliefs were not tested until the first quarter of the twentieth century, when sufficient empirical and experimental evidence had been accumulated to discredit many of these ideas.

The second meaning of "common sense" refers to generally accepted, empirically verified knowledge. In this definition, information derived through common sense represents a body of already investigated data which may be modified as additional evidence is obtained. Thus, the present-day educator may "take it for granted" that quiet children are not always the best-adjusted ones, a position that was not always accepted. Nonetheless, these "common sense" beliefs can be re-examined and reformulated as new information is accumulated.

The Appeal to Authority

The uncritical acceptance of someone else's position has retarded knowledge about education just as it has retarded knowledge in other disciplines. Throughout history, there have been men who were castigated because they

refused to accept the word of some "authority." In the Middle Ages, the teachings of Aristotle and the Church were used as a means of coercing people into "accepting" current religious practices. In more modern times, references to John Dewey or Robert Maynard Hutchins as authorities in education have been used to justify one's own position and to stifle communication and inquiry. When an argument is started with "Dewey said . . .," it is assumed that there is little need to discuss the matter further.

The appeal to authority is not always an unreasonable way of dealing with certain kinds of information. Knowledge has grown too rapidly for anyone to be equally competent in all subjects. Thus, we are forced to accept someone's word if we cannot, either by aptitude or inclination, obtain the evidence for ourselves.

Mander² has proposed four criteria for evaluating the extent to which a statement may be considered authoritative. First, *the individual who is being judged as an expert or specialist should be identifiable*. That is, we must be able to discover the name of the person who has made a statement designed to be authoritative or who has advocated some position. If we want to be precise, we cannot say "Many educators believe . . ." or "Specialists in education claim . . ." without identifying just which educators or specialists we mean. In scientific reports or in theses or dissertations, the author is responsible for providing the reader with a reference to these specialists so that they may be identified and their positions evaluated.

A second criterion suggested by Mander is that *the authority should be recognized as such by members of the profession in which he claims competency*. His membership in professional societies and his reputation among his colleagues may be used as a rough but useful measure of the extent to which the authority's statements have the sanctions of that professional group. To be sure, the danger always exists that personal prejudices may discredit some member of a profession, or that legitimate contributions which are extreme and not well recognized within a profession may be discredited by the general membership. Nonetheless, failure to apply this criterion can lead to an intellectual anarchy in which the persons most capable of evaluating a statement are denied the opportunity to do so.

A third criterion is that *the cited authority should be living*. In ancient Greece, Aristotle could reasonably have been considered an "expert" in matters relating to science and logic; however, he could only be considered an expert in science today if no important or contradictory facts or arguments had come to light since he died. Similarly, it may be quite reasonable to quote John Dewey's position on a number of matters if the purpose is to present historical information. But unless we are certain that no contradictory evidence has come to light since Dewey's death, a reference to him should not

²A. E. Mander, *Logic for the Millions* (New York: Philosophical Library, 1947), pp. 57-60.

be used to "prove a point." Dewey himself might wish to change his position if he were alive to examine new evidence.

The fourth criterion is that *the authority should not be biased*. That is, we need to know whether or not his prejudices, biases, and stereotypes are interfering with his making a clear and rational judgment. Admittedly, such an evaluation is difficult, but difficulty should not prevent us from trying to make these judgments if it is possible to do so.

The Appeal to Intuition and Revelation

Revelations are direct and immediate insights concerning "truth" or "reality"; they are often presumed to originate from God. If these hunches or experiences are believed to have "natural" origins, they are called intuitions.

At one time or another, each of us has probably experienced intuitions or hunches telling us, for example, that we should not trust a person we have just met, or that taking a particular fork in the road when lost would lead us to our destination. These intuitions, derived from past experiences with similar phenomena, may or may not provide us with dependable knowledge concerning the expected or predicted consequences of the decisions we make. The stranger we distrusted may, if given the opportunity, turn out to be a loyal and good friend; the unselected road may have led us home more quickly and safely than the one chosen on the basis of "intuitive" feelings. Intuitive feelings cannot always be trusted without empirically verifying their consequences.

In a sense, the scientific hypothesis or hunch concerning the expected relationships between variables is a sophisticated and highly refined form of intuition. Suppose that a scientist is interested in studying test difficulty levels and their possible effects on student achievement. To investigate this problem, he hypothesizes or suggests that achievement will be facilitated by hard (or easy) examinations. These expectations are derived from his past experiences with similar types of phenomena. In each case, the scientist empirically verifies or disproves his expectations concerning these relationships. In contrast, the appeal to intuition does not necessarily provide for confirmation of one's beliefs.

Revelations are more difficult to analyze than are intuitions, since their source is presumed to originate outside of man's experience. Because the belief in a supernatural existence rests upon faith, the scientist has no way of validating or invalidating revelations unless they contradict empirical experiences. Proclamations that the world will end on a certain day are invalid when that day has passed safely, no matter what the source of the revelation may be. If there is a conflict between a private "revealed truth" and empirically derived knowledge, the strength of the argument will favor empiricism. Even though there is nothing final or absolute about scientific knowledge, the methods employed by scientists involve a set of rigorous criteria for evaluating "truth." Anyone interested in examining a scientific finding can do

so provided he is capable of following the methods used by the scientist. The private nature of a revelation makes it impossible to examine it openly or to duplicate it for further consideration. Thus, while the revelation may have tremendous importance for the person experiencing it, its private nature does not provide a very convincing argument for others to accept whatever "truth" has been revealed.

The Appeal to Reason

Rationalism is the philosophic position that places reason over both revelation and experience. In general, rationalists take the position that concepts such as God, causality, or mathematical "proofs" do not depend upon either experience or revelation but can be proven through rational processes, especially deduction.

Deduction is the process of drawing specific conclusions from premises in a form known as a syllogism. The most common type is the categorical syllogism, which takes the form

All A is B (or, for example, All men are mortal)

C is A (Socrates is a man)

Therefore, C is B (Therefore, Socrates is mortal)

The conclusion that Socrates is mortal is a logical deduction from the major premise that "All men are mortal" and from the minor premise, "Socrates is a man." Since we are setting equalities equal to each other ($A = B$; $C = A$; therefore, $C = B$), the conclusion is valid.

Deduction is not an unreasonable way of deriving knowledge. Mathematical reasoning, for example, is dependent upon rules of logic which specify the definitions of various terms and their interrelationships. The scientist employs these rules and definitions to help him derive relationships and specific conclusions. For example, it is a simple matter to demonstrate mathematically that adding a constant to a series of scores will augment the mean of the original series by the value of the constant. This can be accomplished by 1) defining each term needed and 2) performing operations in accordance with rules that do not violate logical principles.

It is one thing to consider deduction as an auxiliary and useful methodology; it is quite another matter to believe that it is the primary method for deriving all "truth." In deductive reasoning, the conclusions may tell us nothing about "reality," since we may have major or minor premises which themselves do not correspond with the "real" world. Thus, we are on safe ground when we state that "All men are mortal," since "mortal" and "men" are semantic equivalents. But nothing prevents us from stating "All men are elephants" and "Socrates is a man." The conclusion must then be drawn that "Socrates is an elephant." Note that the rules of logic do not specify what the empirical relationships must be between men, elephants, and Socrates. Conclusions may be logically derived although premises are incorrect.

SCIENCE AS A SOURCE OF KNOWLEDGE

This text takes the position that science provides the most reliable means for obtaining knowledge about empirical relationships and that education can be studied scientifically. The acceptance of this position does not necessarily exclude appeals to common sense, authority, intuition, or reason as long as these appeals do not violate those characteristics of science we shall consider later in this chapter.

The term *science* has been used in so many different ways that we must consider its meaning in greater detail. In attempting to win an argument, for example, it might be argued that the opposition does not really have the facts correct because "Science has proven . . ." or because "It is a scientific fact that . . ." The argument that begins "Science has proven . . ." assumes that counter-arguments are invalid. To use science in this way is to establish its principles as authoritative without the necessity of providing evidence for one's beliefs. To assume that scientific "truths" are not to be questioned is to misrepresent the very meaning and purpose of science.

A second meaning of the word *science* is implied in the expression, "He has that down to a science." Used in this sense, science refers to some skill which has been perfected to a high degree of sophistication; it thus refers to and implies exactness, precision, and perfection. The layman expects science to allow for no error at all. But science, of course, cannot be independent of scientists, and wherever human beings are involved, some degree of error and inexactness is to be expected. We will examine this concept in more detail later in this chapter.

A third meaning of *science* is implied in such expressions as "scientific barbering" or "scientific housecleaning methods." In these examples, science is used as an honorific designed to raise the prestige value of these activities. Just how "scientific barbering" differs from the ordinary kind is never really considered seriously by those employing the term. The intent is to impress rather than to explain.

A fourth meaning of *science* refers to an organized body of knowledge, such as library science or political science. If this organized body of knowledge is based primarily on deductive rules it is called a *formal science*, to distinguish it from those areas of knowledge that depend primarily upon empiricism and induction. These are called *empirical sciences*. Considered from this point of view, mathematics, logic, and library science are primarily formal sciences, whereas chemistry, psychology, and education are primarily empirical sciences.

There are a number of difficulties in defining science as a body of knowledge. A telephone book, for example, is a highly organized body of knowledge, but it is hardly a branch of science. Alchemy and astrology were also well organized but are not sciences as the term is used today. Rather they form a body of *pseudosciences*, since the methods they employ do not correspond to those required for modern science.

Most scientists today consider science from a fifth point of view, that is, as a method or attitude. As a method, empirical sciences consist of controlled observations combined to form "an interconnected series of conceptual schemes . . . fruitful of further experimentation and observation."³ Observations are controlled by eliminating or reducing the effects of irrelevant or extraneous variables by means of apparatus, tests, examinations, instruments, or specialized techniques. These observations are combined to form *theoretical systems* that are useful in explaining and predicting the relationships between the observations.

As an attitude, science consists of a willingness to examine and, if need be, to modify or repudiate one's own beliefs, ideas, or methods as new or more reliable evidence is obtained. The attitude is one that encourages the scientist to control and reduce biases, to be as precise as possible, to verify or disprove his beliefs empirically, and to organize empirically obtained data into a meaningful body of theoretical knowledge.

There is no necessary conflict between thinking of science as a method and thinking of it as an attitude, as long as method is defined broadly. Certainly we do not mean that astronomers, chemists, and psychologists use the same specific techniques for obtaining data; we do mean that all sciences are concerned with obtaining the most reliable empirical observations possible, even though the specific techniques for obtaining this information may differ widely.

CHARACTERISTICS OF SCIENCE

Reduction and Control of Bias

The scientist is a human being and, like others, he is not without prejudices, biases, and misinformation. The conception of the scientist as a cold, detached, and impersonal individual reflects a popular but misleading stereotype. In one study which explored the attitudes of high school seniors toward science and scientists, the investigator found that scientists were considered to be "longhairs," "shy, lonely individuals," "odd," "eggheads," and "communistic." They "have physical deformities which render them unfit for other work," "cannot have a normal family life," and are "unsociable."⁴

This stereotype of the scientist is not unlike the character of Dr. Frankenstein as displayed in the films! It may very well be that *some* scientists are odd and unsociable, but these are the exceptions, not the general rule.⁵ A number

³James Conant, *Science and Common Sense* (New Haven: Yale University Press, 1951), p. 25.

⁴Hugh Allen, Jr., *Attitudes of Certain High School Seniors Toward Science and Scientific Careers* (New York: Bureau of Publications, Teachers College, Columbia University Press, 1959), pp. 14-15.

⁵Paul F. Brandwein, "Creativity and Personality in the Scientist," *Rethinking Science Education, 59th NSSE Yearbook*, Part 1 (Chicago: University of Chicago Press, 1960), pp. 62-81.

of early writers such as Lombroso,⁶ Kretschmer,⁷ and Lange-Eichbaum⁸ helped popularize this misconception of the scientist and made it difficult to consider scientists favorably. This stereotype may be responsible for discouraging students from becoming scientists or from engaging in scientific pursuits.

The emotional sterility often inappropriately demanded of the scientist would no doubt interfere with some aspects of scientific progress. Nothing would be more boring than having to compute pages of statistics, score innumerable tests, or read countless articles without a driving and compelling interest either for this type of work or for the end product of these activities. The very selection of a research problem is a value judgment made by the scientist, and his interest in doing one type of research rather than another reflects this personal preference.

In another sense, personal preferences may interfere with scientific pursuits. We can understand the scientist's preference for one type of research over another and for one procedure over another, but it is quite another matter if he allows his personal values and biases to interfere with the methods he employs in solving some problem, or if he allows his prejudices to interfere with the conclusions he draws from his data. It is one thing when the scientist's preferences lead him to work on problems of compelling personal interest; it is quite another matter when he systematically excludes evidence that is contrary to his personal beliefs.

The Quest for Precision

In the high school science survey referred to earlier,⁹ almost half of the seniors agreed with the statement: "Scientific findings always lead to final truths." What these seniors believed was that science cannot err and that "final truths" are obtainable.

Science demands as much rigor and exactness as can be mustered, but this demand should not be confused with a search for absolute or irrefutable "truths." Man can depend only upon himself, his methods, and his tools in his attempt to understand his environment; he has no appeal to higher "authorities" to validate his methods or conclusions.

Our culture contains many examples which point out man's quest for absolute knowledge and certainty. The young child looks to his parents as the source of irrefutable "truth." Later, the teacher and the "expert" are considered the authorities. It is little wonder, then, that graduate students feel it is time they are admitted to the "inner circles." What they fail to realize is that there are no absolute or final answers in science.

⁶Cesare Lombroso, *The Man of Genius* (London: Scott, 1896).

⁷Ernst Kretschmer, *The Psychology of Men of Genius* (New York: Harcourt, Brace & World, Inc., 1931).

⁸Wilhelm Lange-Eichbaum, *The Problem of Genius* (London: Routledge & Kegan Paul Ltd., 1931).

⁹Allen, *Attitudes*, p. 24.

In an ingenious demonstration, Bernatowicz and Kay pointed out man's search for certainty by constructing a Black Box, containing objects of various sizes and shapes. To determine the contents of the Black Box, their students were provided with apparatus such as thermometers, rulers, scales, and probes, which they could insert in small holes located on the walls of the box. In this demonstration, the students were asked to discover as much about the contents of the box as they could without damaging it. Bernatowicz and Kay concluded:

The most important lesson comes at the end of the critique when the students invariably ask that the boxes be opened to disclose what is "really" inside. Here we emphasize that it is futile to hope for final answers in science. In principle, the Black Box represents any problem in science. Each problem is attacked by means of observations which lead to interpretations To ask what is "really" inside is a futile question unless one has some definition by which he will recognize reality when he meets it.

The matter of wanting the Black Box opened strikes us as a manifestation of a childlike desire for certainty. It betrays an inability to recognize that one's observations of the physical universe should be the authority for his conclusions about it, and that in such matters one can never hope to *know* in some absolute sense. Opening the Black Box, "looking it up in the book," expecting some higher authority to grade us right or wrong are all alike in this regard. They obscure the individual's right and responsibility to gather evidence, to formulate opinions, and to live by the consequences of his own judgment. . . .¹⁰

Bernatowicz and Kay were not suggesting that all observations are equally competent or valid. Rather, we strive for as much precision and accuracy as our tools, resources, knowledge, and intelligence can give us, while recognizing the tentative nature of all findings and conclusions.

Verification

The scientist reports his observations in such a way that not only he, but other investigators, can confirm or reject them. Thus he recognizes the fallibility of his own findings, and also recognizes that other individuals have the right to question and investigate these findings. This "self-corrective" function is perhaps the single most important aspect of science. In essence it means that the scientist has an *obligation* to present his evidence in a manner that allows it to be confirmed or rejected by other investigators. Science does not, or at least should not, have secret or sacred information. Indeed, scientific research depends upon other investigators checking on reported evidence and verifying or disproving it.

The scientist can only work in areas in which verification is possible. Thus, on a rather simple level, it may be verified that a certain flower has five petals, that a given leaf is green, or that the sun is shining. On a more complex

¹⁰Albert J. Bernatowicz and E. Alison Kay, "Scientific Attitudes, Certainty, and the Black Box," *Journal of General Education*, XIII, No. 1 (1961), 25-29.

level, it may be verified that objects drop with constant acceleration or that planets closer to the sun travel at faster rates than those farther away. These more complex facts may be verified by taking careful observations and measurements.

However, the statement "Honesty is the best policy" cannot be verified empirically. It is a value judgment which rests upon certain philosophical assumptions concerning the nature of man, the purposes of society, and the meanings of such words as *honesty*, *best*, and *policy*. Empirical observations can neither confirm nor invalidate value judgments since the criteria for evaluating them differ from those used to evaluate empirical observations. Value judgments can be tested by employing hedonistic, utilitarian, or perfectionistic criteria. Thus, the hedonist might construe *honesty* to mean "honesty to one's self-interests"; to the utilitarian, it might be interpreted in light of its possible effects on mankind in general; and the perfectionist might conceive of it in terms of some ideal Utopian state.

In contrast, empirical statements can be evaluated directly through observation or indirectly by setting up consequences which themselves can be directly observed. Thus, the fact that it is now daytime can be confirmed by looking outside or examining clocks graduated on a 24-hour basis; it can also be confirmed indirectly by testing the consequences of its being daytime. We might argue, for example, that if it is daytime the stores, shops, and schools will be open. If they are, this is some evidence in our favor, but it is by no means conclusive, since stores, shops, and schools do stay open in the evenings from time to time and may be closed during the day on weekends or holidays. However, as we test the consequences of more and more propositions concerning the time of day, we can increase the likelihood that we will make a correct judgment. We shall return to this point in Chapter Five, when we discuss the role of the hypothesis in scientific research.

Empiricism

Empiricism takes the position that observation and experience are the prime means of obtaining knowledge; it is the basic method of the sciences. At one time empiricists rejected all knowledge not derived from sensory experiences, including inferences drawn from deduction, intuition and revelation, and theory. However, a less extreme empiricism is generally accepted by a majority of scientists today. This more moderate position encourages the development of theory and allows for deduction, but denies intuition and revelation as sources of knowledge. The empiricist would insist, however, that theories yield consequences which may be empirically verified.

The basic form of reasoning used by empiricists is *induction*, which draws general conclusions from specific observations. Inductive reasoning follows the general form

Observation 1

Observation 2

Observation 3

.

.

.

Observation n

Generalization from observations 1 . . . n

In induction, observations 1 through n are made; from them the scientist attempts to induce a general principle. Whether or not the generalization is valid depends in part on the form or-type of induction being employed. If we have knowledge of all observations comprising some universal or totality, we can summarize that which holds true for all members of the set. Thus, if we know that teachers 1 through n in a given school district are females, we can claim that all are females, since we have complete knowledge of all elements entering into the set. Where this is so, the process is called *perfect induction* or *induction by complete enumeration*. Essentially, perfect induction is a statement summarizing what is known about all members of any given totality.

Usually we do not have complete knowledge of all elements comprising any given set, either because the elements are infinite in number or because it may be extremely difficult or expensive to examine them all. In either case, the scientist works by selecting samples to represent all of the elements in any given universe (see Chapter Six) in the hope of drawing conclusions from the sample that can be generalized to the universe as a whole. If sampling is involved, the form of induction is called *imperfect*, to indicate that conclusions do not necessarily or logically follow from the enumeration of limited numbers of observations. However, the scientist often can estimate the amount of error there will be when he generalizes from the sample to the universe.

There are a number of criteria that can be used to evaluate imperfect inductions. In general, we are likely to make fewer errors if we can examine a large proportion of the elements in the universe. For example, we now have accumulated thousands of observations verifying that the sun rises in the east. The first criterion argues that the more of these observations we can get, the more certain we can be that the sun will always rise in the same direction. However, this criterion, while necessary, is not sufficient because we also need to know the number of negative instances, or times the sun failed to rise in the predicted direction. Here again, we have no negative instances and the generalization seems even more plausible. Still, for the third criterion, we must know how capricious the phenomena are which are under investigation.

Some phenomena, such as human behavior, may vary considerably, while others obey laws that are unlikely to change. The laws of motion that govern planetary movements are highly stable, and we can predict that, as long as the solar system is not disturbed, the sun will always rise in the east. It is important to note that this conclusion was not drawn only from the fact that the sun has always risen in the east: the number of negative instances and the variable nature of the phenomenon under investigation also had to be considered.

We pointed out earlier that scientists depend upon rational as well as empirical methods. Mathematics, for example, is almost entirely deductive; yet, because there is an *isomorphic* relationship between it and the empirical world, the two can be extremely useful to each other. Mathematical models can help predict empirical relationships and these, in turn, can provide an empirical check on the model. Empiricism and deduction, then, can be mutually rewarding.

Theory Construction

The word *theory* has been used in a number of different ways. In one sense, theory might be considered to be at one end of a continuum with "practical" matters at the other end. In this way, theory becomes a form of contemplation,¹¹ whereas "practical" matters involve actual manipulation, or "doing." However, since science always involves thinking and contemplation, it is all theoretical to some extent.

In another and more specious sense, theory may be considered as contemplation which cannot be translated into any practical or useful activity. Campbell has stated, "The idea that there are propositions 'true in theory, but false in practice' has its foundation only in the incompetence of the uninitiated to understand theory, and in their habit of applying propositions to circumstances entirely foreign to the theory. To those who have not the power to think, theory will always be dangerous."¹²

A theory is a unified system of principles, definitions, postulates, and observations organized in such a way as to most simply explain the interrelationships between variables. Since nature does not organize and interrelate objects, events, or methods for us, it is up to man to organize nature in the manner most meaningful for the purposes at hand. Thus, while it might be quite reasonable for some purposes to classify all objects as large or small, it may be less meaningful to do so when constructing kingdoms in biology.

Cohen and Nagel have emphasized that neither isolated facts nor organized information alone can be considered as science.

¹¹Norman R. Campbell, *Foundations of Science: The Philosophy of Theory and Experiment* (New York: Dover Publications, Inc., 1957) p. 120. The quotation was originally published in Norman R. Campbell, *Physics: The Elements* and is reprinted here by permission of Cambridge University Press.

¹²Campbell, *Foundations*, p. 121.

Information, no matter how reliable or extensive, which consists of a set of isolated propositions is not science. A telephone book, a dictionary, a cook-book, or a well-ordered catalogue of goods sold in a general store may contain knowledge, organized in some convenient order, but we do not regard these as works of science. Science requires that our propositions form a logical system.¹³

Thus, simply to describe the events which occur in large numbers of classrooms is not science, even though such information has been gathered by the most careful observations. Nor can well-organized class rosters be considered as science. They simply exist as isolated examples of internally organized information that does not help to predict or explain the relationships between variables. On the other hand, the atomic theory consists of definitions, assumptions, and empirically confirmed facts which are interrelated to provide an explanation of natural events.

The Role of Induction in Theory Development

The role induction plays in developing a theory may be shown by an example. Suppose that the following "facts" have been verified, reconfirmed, and are universally "accepted":

1. There is a positive relationship between the amount of material which has been forgotten and the amount of time which has passed since the material was originally learned.
2. The beginning and end of a memorized list will be retained to a greater extent than will material in its center.
3. After sleeping, pupils can recall data they could not recall the night before (the "reminiscence" effect).
4. The more material learned, the more forgetting is likely to occur.

The scientist attempts to develop or induce from particular observations a theory or generalization which will explain all four facts. Suppose that the following statements are considered as possible explanations:

- a. *Theory of Disuse*: Forgetting occurs because material which has not been practiced degenerates.
- b. *Psychoanalytic Theory*: Forgetting occurs because unpleasant memories are repressed.
- c. *Interference Theory*: Material learned early in a sequence interferes with the learning of later material (*proactive inhibition*), and material learned later in a sequence tends to interfere with previously learned material (*retroactive inhibition*).

To what extent does each of the three theories explain all of the facts? The Theory of Disuse can account for the apparent "decay" of memory over time, but it does not explain facts two or four and is in direct contradiction to fact three.

¹³Morris R. Cohen and Ernest Nagel, *An Introduction to Logic and Scientific Methods* (New York: Harcourt, Brace & World, Inc., 1934), p. 129.

The Psychoanalytic Theory may partially account for the so-called reminiscence effect, but it does not explain the other facts. The only theory which seems to account for all of the facts is the Interference Theory. It explains fact one by asserting that more interference is possible over long periods of time than over shorter intervals. It explains fact two on the grounds that the terminal part of any sequence to be memorized can only be interfered with by preceding parts (i.e., affected proactively), while the beginning of the sequence can only be affected by material which is learned later on (affected retroactively). However, the material in the center of the sequence is affected by preceding as well as by following events and thus becomes most interfered with (affected both proactively and retroactively). The Interference Theory only partially explains fact three by postulating that during sleep there is little chance for interference to occur. Fact four is explained by the greater amount of interference possible with a larger amount of material to be learned.¹⁴ In the process of induction, one ordinarily tries to account for as many facts as possible, by developing the fewest number of theories or explanatory generalizations. The most useful theories are those which most adequately and simply explain the greatest number of relevant facts.

The Relationship Between Theories and Facts

The researcher developing a theory must be willing to modify it to account for discrepant facts; the opposite process, modifying facts to conform to a pet theory, has been called "Maier's Law."¹⁵ It states: "If the facts do not conform to the theory, they must be disposed of."¹⁶ Maier has indicated a number of ways facts have been ignored to perpetuate a poor theory. One way is to give the facts a new name.

Giving disturbing facts a name is almost as good as explaining them because a name supplies a useful answer to inquisitive people. For example, a lecturer in describing the habits of people living near the North Pole told his audience how children ate blubber as if it were a delicacy. Later a questioner asked the speaker why these children liked a food that would not be attractive to children living here. The lecturer replied that this was so because the children were Eskimos. The questioner replied "Oh, I see" and was satisfied. In a similar manner the word "catharsis" explains why we feel better after expressing pentup feelings.¹⁷

Facts may also be disregarded in another way. According to Maier, the most efficient way of eliminating facts that do not correspond with theory is to fail to report them.

¹⁴The reader should recognize that these "facts" and "theories" were oversimplified to make the process of induction clear. As such, the example is illustrative but does not accurately correspond to the actual facts or theories.

¹⁵N. R. F. Maier, "Maier's Law," *American Psychologist*, XV, No. 3 (1960), 209-212.

¹⁶Maier, "Maier's Law," p. 208.

¹⁷Maier, "Maier's Law," p. 209.

Naturally it takes a critical individual to determine what is worth reporting and what is irrelevant. For example, one researcher reported that his experiment on delayed reaction was conducted on the third floor of the building but did not tell how many tests he ran in a day. Later it was found that the number of tests per day determined the length of a delay, while the floor used was not important. Since selection is always with us, what better aid is there for the selecting of facts than a good theory.¹⁸

The relationship of fact to theory has not always been well understood by many researchers. The student should keep in mind how theory may be used to clarify facts and how facts may be used to validate theory. Van Dalen¹⁹ has indicated a number of ways in which theory and facts are interrelated:

1. *Theory defines the relevancy of facts.* Not all facts are of equal value. Some facts, although highly dependable, may not be related to anything of immediate importance. Thus we could collect facts about the sizes of foreheads of elementary school children or try to determine the number of songs seniors can learn over a given period of time. But these facts, once obtained, have limited applicability. Having a theoretical structure helps determine which kinds of facts are needed to validate or invalidate a given theory.

2. *Theory develops systems of classification and a structure of concepts.* Since nature does not provide us with a ready-made system of classifying objects, events, or methods, it is up to man to organize facts in such a way that they become useful. A theoretical structure helps the scientist—whether he is a physicist or an educator—to organize his facts by relating them to a broader concept of theory.

3. *Theory summarizes facts.* It may be extremely difficult for the scientist to retain large numbers of seemingly unrelated facts. However, as a theory is developed which encompasses these facts, he can summarize large masses of data by subsuming them into a meaningful structure. Thus, one theory may explain a large number of facts.

4. *Theory predicts facts.* One of the major purposes of theory is to predict events in the absence of direct evidence. A novice teacher, for example, may have rather specific questions concerning the behavior of a given child in his classroom. He may want to know why the child is continually hitting other children in the room. Without the benefit of theory he would have to investigate in detail the characteristics of every child before he would be in a position to discuss aggressiveness in children. Once a theory has been established and confirmed in a number of different situations, it enables him to predict within the limits of probability what will happen when certain events

¹⁸Maier, "Maier's Law," p. 210.

¹⁹D. B. Van Dalen, "The Relationship of Fact and Theory in Research," *Educational Administration and Supervision*, XLV, No. 5 (1959), 271-274.

occur. Thus, a psychoanalytic theory might suggest that aggression is displaced hostility; another theory may suggest that it is related to the amount of frustration experienced. No matter what the specific form of the theory, it allows the teacher to predict what the child will do, and may offer an explanation for his behavior.

5. *Theory points out needs for further research.* A fruitful source of research topics can be found by attempting to confirm or refute theories already proposed. All theories—and especially those in the social sciences—are tentative formulations which can be changed to conform to known facts. By examining areas where there is a lack of information, the researcher may find useful ideas.

An example may help to clarify this process. Suppose one does accept the Interference Theory of forgetting (see page 13). One way of selecting a research topic is to test the implications of the theory itself. For example, the Interference Theory states that retention is a function of the amount of interference. Suppose that this theory is tentatively "accepted." What should we then be able to predict concerning forgetting? We might predict the very "facts" listed on page 13. These facts then become a possible topic for investigation. Each theory should be able to generate hypotheses which can be experimentally tested to clarify, expand upon, and delimit various aspects of the theory itself.

So far we have indicated ways in which theories contribute to knowledge. It is also true, however, that facts contribute to theory.²⁰ First, facts stimulate theory construction. As more and more facts are added to any body of knowledge, the need to organize and explain these facts increases correspondingly. As this organizational process continues, it forms the nucleus for the development of theory. Thus, it is quite reasonable to select research topics to obtain facts, which in turn develop some body of theory.

It is also reasonable to select a research topic which has as its intent the testing of the implications of a theory. For example, the theory behind progressive education assumed that the "problems approach" to the teaching of arithmetic was "superior" to more deductive approaches. A study by Sax and Ottina²¹ indicated that children from progressive schools did no better than their matched counterparts in traditional schools on tests involving arithmetic meanings until they reached the seventh and eighth grades; then, significant differences were found in favor of children who attended progressive schools. These facts help to redefine the scope and extent of the theory of progressive education.

Simplicity

The so-called *Law of Parsimony* states that of several equally suitable theories, science will tentatively "accept" the simplest. The Law of Parsi-

²⁰Van Dalen, "The Relationship of Fact and Theory in Research," pp. 273-274.

²¹Gilbert Sax and John R. Ottina, "The Arithmetic Achievement of Pupils Differing in School Experience," *California Journal of Educational Research*, IX, No. 1 (1958), 15-19.

mony, or the Principle of Occam's Razor (named after William of Occam, who suggested "cutting away" unnecessary explanations), is not so much a law as it is a suggestion to the scientist to make his explanations and theories as simple as he can. It advises the researcher to conceptualize his propositions in the simplest terms possible; of course it does not suggest that he oversimplify complex and relevant data.

An example of the use of the Principle of Occam's Razor may help to clarify it. Suppose that a researcher makes the following "discoveries":

1. From 7:00 A.M. to 8:00 A.M. there are no students at A High School.
2. The number of students remains constant from 8:00 A.M. until 3:00 P.M. each day from Monday through Friday.
3. From 3:00 P.M. until 4:00 P.M. the number of students on campus continually diminishes.
4. From 4:00 P.M. until 8:00 A.M. the following day there are no students on campus.

Assuming that these facts may be verified, what explanations or theories can account for all of them? One possible theory is that policemen are posted around the campus keeping students out until 8:00 A.M. and not letting them leave until 4:00 P.M. Another plausible theory, which also accounts for all of the facts but is somewhat simpler, is that classes do not begin until approximately 8:00 A.M. and that school ends at 3:00 P.M. A third theory might be that students have to help their parents at home until school begins at 8:00 A.M. and are required by their parents to be home to do chores no later than 4:00 P.M.

Each of these three theories accounts for all of the facts. The first theory, however, does so only by adding the role of a policeman. Like the third theory, it requires additional propositions to explain the phenomenon adequately. The second theory, however, explains all of the facts without elaborate additions. Therefore, it is simpler and needs no additional postulates to account for all of the events.

One caution is in order at this point. The Law of Parsimony does not state that only simple theories can be accepted, nor does it state that complex explanations are inappropriate. In addition, acceptance of an explanation does not necessarily indicate complete agreement. Adherence to the principle of parsimony does, however, assume a hierarchy of tentative explanations, each of which must be tested in greater detail. Thus, the researcher "accepts" an explanation only in the sense that he plans to investigate it first.

SOME ASSUMPTIONS MADE IN SCIENCE

Determinism

Determinism is the philosophical assumption that the universe is orderly, lawful, and predictable. It is often contrasted with *indeterminism*, or *volunta-*

rism,²² which postulates that man is not determined by his environment but is free to determine his own goals and behavior. A voluntaristic universe could not be studied because there would be no predictability, no stability, no order.

In support of determinism, B. F. Skinner, the Harvard behaviorist, has stated:

Science attempts to discover order, to show that certain events stand in lawful relations to other events. No practical technology can be based upon science until such relations have been discovered. But order is not only a possible end product; it is a working assumption which must be adopted at the very start. We cannot apply the methods of science to a subject matter which is assumed to move about capriciously. Science not only describes, it predicts. It deals not only with the past but with the future. Nor is prediction the last word: To the extent that relevant conditions can be altered, or otherwise controlled, the future can be controlled. If we are to use the methods of science in the field of human affairs, we must assume that behavior is lawful and determined. We must expect to discover that what a man does is the result of specifiable conditions and that once these conditions have been discovered, we can anticipate and to some extent determine his actions.²³

Educational research presupposes that behavior is predictable and that events act lawfully in relation to one another. If this concept of determinism cannot be accepted, no science of education is possible.

A number of arguments have been raised against determinism. One can argue, for example, that human beings are simply too complex to be studied scientifically. To this charge we can answer that modern experimental designs and the use of high-speed electronic computers have added greatly to the researcher's ability to handle and control highly complex problems. To be sure, we know little about the behavior of children in classrooms, but there is no need to assume from this that the information is unobtainable.

Two other points should be made. First, the concept of determinism does not (and indeed could not) require us to predict with certainty the results of any given action or behavior. The extent to which an event occurs or does not occur is defined by its probability of occurring: some events have near-perfect probabilities, others have extremely low ones. Events do not ordinarily occur as *either/or* propositions but in degrees of certainty ranging from zero to unity.

Second, determinism should not be confused with *fatalism*. As Grunbaum states:

The fatalist says that regardless of what we do, the outcome will be the same. By contrast, the determinist says that *if* we do such and such, *then* the effect will be thus and so. The fatalist thinks that if you go into combat, and

²²Gustav Bergmann, *Philosophy of Science* (Madison: The University of Wisconsin Press, 1957), p. 106.

²³B. F. Skinner, *Science and Human Behavior* (New York: The Macmillan Company, 1953), p. 6.

if "some bullet has your name on it," you will be killed no matter what you do. . . . The determinist maintains that a person will die on a certain day only if the conditions which lead to death materialize for that person on that day, as indeed they do at some time for each of us. Unlike fatalism, determinism allows causal efficacy to human actions.²⁴

Nature of Reality

The very language we use to convey our thoughts and ideas may presume a knowledge of "reality." Thus, some scientists might speak of "body and mind," "method and content," or "natural and supernatural," thereby assuming that the universe is composed of two separate and distinct types of realities, one having to do with substance and the other with "mind." Such persons (called *dualists*) conceive of "ultimate reality" as composed of both "physical" and "mental" entities.

Others believe that "ultimate reality" is composed only of "real, substantial entities, existing in themselves and ordered to one another by extramental activities."²⁵ These persons, called *realists*, conceive of a purely physical universe which has an existence independent of man. Still others may consider ideas as the "ultimate reality" and would deny a completely physical universe; to *philosophical idealists*, reality consists of ideas.

Metaphysics is that branch of philosophy concerned with the systematic interpretation of reality. Although the study of metaphysics is not itself an empirical or scientific enterprise, knowledge of one's assumptions concerning the nature of the universe often directly benefits the scientist. These assumptions, if not recognized and made explicit, are likely to find their way into scientific reports.

Consider, for example, an investigator who states, "Classroom learning is both intellectual and emotional." He has assumed that there is a difference between intellectual and emotional learning and that both types exist. This is not a question which the scientist can resolve, because there is no possible way of verifying what "ultimate reality" is composed of. However, unless the student is aware of the metaphysical assumptions he is making when he uses words such as "absolute" or "relative" or makes distinctions between "natural" and "supernatural," he is likely to be led astray by his own language.

The importance of recognizing assumptions can be demonstrated by another example. Take the case of the investigator who states: "The real problem of underachievers is their lack of confidence." In this statement, he has made a number of assumptions. First, his use of the word "real" leads the reader to believe that there are at least two types of problems, one type which is

²⁴Adolf Grunbaum, "Causality and the Science of Human Behavior," *American Scientist*, XL, No. 4 (1952), 671.

²⁵John Wild, "Education and the Human Society: A Realistic View," *Modern Philosophies and Education*, 54th NSSE Yearbook, Part 1 (Chicago: University of Chicago Press, 1955), p. 17.

"real," "true," and "absolute," whose answers are discoverable if only we apply the correct methods, and another type which somehow is "false," or "incorrect." However, there is no way to verify what a "real" problem is, since we have no appeals for empirical matters beyond ourselves.

A second assumption concerns the meaning of underachievement. The word "underachievers" assumes that there is a group of students who are achieving, or "working up to their ability." Yet, on a moment's reflection, it seems clear that we have no measures of a student's *ability* obtained independently of his *achievement*.

It is the responsibility of the researcher to recognize the assumptions he makes. Research papers sometimes contain a section titled "Assumptions," which forces the investigator to recognize and make explicit those beliefs he is asking the reader to take for granted.

SOME POSSIBLE OBJECTIONS TO A SCIENCE OF EDUCATION

The Problems of Applying Scientific Methods to the Social Sciences

Objections have been raised to the consideration of education as a scientific enterprise. The most persistently voiced objection is that scientific methods, although they may be of value to the natural sciences, are inappropriate to the social sciences. A proponent of this position has stated:

If it is true that the truth can be discovered only in the laboratory, then we can know very little indeed about education; for we cannot know even whether the statement is true that truth can be discovered only in the laboratory. . . . What is called social science cannot tell us what kind of society we ought to aim at. It is doubtful whether it can even tell what the consequences of a given social policy will be. The reason, again, is the enormous number of variables that enter into any social situation.²⁶

We can agree with the statement that the social sciences, particularly education, face innumerable difficulties in their attempts to solve persisting social problems, without accepting the premise that this knowledge is unattainable. If we reject scientific methods of solving human problems, we should be prepared to substitute some other method. Yet we know of none more rigorous than science in its approach or more likely to yield reliable empirical knowledge.

Not all "truth" can be discovered in the laboratory, and the social sciences "cannot tell us what kind of society we ought to aim at." Nevertheless, attempts to define the objectives of education that do not account for the psychological characteristics of children are likely to fail. These character-

²⁶Robert M. Hutchins, *The Conflict in Education in a Democratic Society* (New York: Harper & Row, Publishers, 1953), pp. 78-79.

there could be an art of teaching, if art is used to mean the subtle practices which "successful" teachers use. But to consider teaching as an art does not eliminate education from being a science.

DESCRIPTIVE AND EXPLANATORY LEVELS OF SCIENCE

Descriptive Level of Science

One purpose of science is to explain events or behavior. In any scientific undertaking, however, description and careful observation ordinarily precede explanation of the events.

When the scientist describes what is going on in a classroom, he tries to point out *essential* characteristics and details of pupil-teacher behavior and of the classroom environment. This description may be *molecular*, noting the smallest details (such as physiological reactions or movements in response to specific stimuli), or it may be *molar*, describing larger and more encompassing acts (such as students raising their hands when the teacher asks a question).

For example, suppose that we wish to describe the behavior of a child who is subjected to various degrees of frustration produced by withholding a variety of toys. At a molecular level, we might wish to measure such physiological reactions as blood pressure or galvanic skin responses, or we might be concerned with the specific *movements* which the child or the teacher make in response to various types of stimuli. If the intent of the research is to investigate physiological reactions to frustration, a molecular approach may be quite reasonable. If, however, we want to describe to teachers how children respond when frustrated, the description may consist of larger units of behavior such as "John cried" or "The children became aggressive." This more molar level of description would be sufficient.

There are consequences of becoming overly molecular or overly molar in describing behavior. An overly molecular description may provide more information than is needed to understand the problem. It requires an excessive amount of work accumulating data which may have little potential use. This would be similar to teaching pupils the concept of a book by describing every typographical mark contained in one. An overly molar description, on the other hand, may be vague. A statement such as "A book is composed of pages" does not describe a book with sufficient clarity to be useful.

Description, whether molar or molecular, is always selective. The researcher must choose certain aspects of his environment to observe and disregard others. To a large extent his training should determine what will be selected for observation and what will be disregarded. To an untrained observer, an extremely obedient and quiet boy may appear to be the very epitome of the "good child." To psychologically perceptive teachers, however, his behavior may be symptomatic of severe emotional problems. In a classroom, or in

mathematical models can be used to represent many empirical relationships. If the model is a physical object (i.e., a picture, photograph, or miniature replica), it is called an *iconic* or *replica* model; if the model does not physically resemble its prototype but is used within a deductive framework to point out relationships, it is called a *symbolic* model.

Models need be no more complex than their purpose demands. Political maps, for example, can be useful even though they may not be perfectly accurate; a model of the atom as a miniature solar system has proven itself useful even though the correspondence between them is by no means perfect. As long as we can reproduce the essential characteristics of a phenomenon, the model can be useful. What these essential characteristics are, of course, depends upon our purpose in constructing or developing the model. If we wish to predict the differential effects of stress on a new type of airplane wing, the replica model should behave in the same way as the "real" one, and we may be willing to disregard such variables as color or comfort if they are unrelated to our main purpose. Symbolic models, such as those based upon the normal curve or upon the algebra of straight lines, have relevance because they are isomorphic with empirical relationships.

Models help to explain phenomena by pointing out the essential similarities and differences between the model and its prototype in much the same way that analogies are used to explain complex ideas or relationships. Thus, for example, electronic computers have been used to form models for the human brain, to simulate large-scale management problems involving the deployment of personnel and equipment, and to estimate the consequences of decisions made by company executives. By using models to test hypotheses, savings in time, effort, and money can be realized.

Explanatory models are of four different types.²⁹ The *deductive model* takes the form of logical deduction. Given certain premises, certain conclusions necessarily must follow. If it is true that $A = B$ and $B = C$, then A must be equal to C . The conclusion follows from a knowledge of the logical relationship between the three variables. Similarly, we can argue that if $5^2 = 25$, then $\sqrt{25} = 5$, and $25^{1/2} = 5$. It should be clear from our earlier discussion of deduction that the deductive model may be operative whether or not the premises are valid.

A second type of explanatory model involves *probability*. John bites his nails, is nervous, has difficulty sleeping, perspires excessively, and cries a great deal; a likely explanation for his behavior is that he belongs to a general class of individuals who, for one reason or another, are unable to adjust to the stress of modern living. With a probability model, John becomes a member of a general class of individuals (neurotics) about whom we have some information of a statistical nature. Thus, from empirical studies, we may

²⁹Ernest Nagel, *The Structure of Science* (New York: Harcourt, Brace & World, Inc., 1961).

a. Common sense may refer to "generally accepted but incorrect information," or to "generally accepted and empirically verified knowledge." In the first definition the term is used to avoid presenting evidence for beliefs; the second use of the term, while more defensible, is too often used as a rationalization to end discussion.

b. The appeal to authority is an attempt to "take someone else's word for it," rather than to investigate matters as carefully as possible for oneself. Four criteria for judging authorities were suggested: 1) the authority must be identifiable; 2) he must be recognized as such by members of the profession in which competency is claimed; 3) he must be living, or no new and relevant facts should have come to light since his death; and 4) he must not be biased.

c. Intuitive beliefs are "hunches" or "insights" into the relationships between variables which suggest a course of action and which are derived from one's past experiences with similar phenomena. If put to the empirical test, intuitions can be verified or disproved, providing a check upon one's beliefs. In this way intuition is similar to the scientific hypothesis. Both test empirically the expected relationships between variables. If intuitive beliefs or feelings are not verified or disproved empirically, they are not validated and can provide no justifiable basis for action but that of convenience.

Revelations are beliefs or feelings claimed to be derived "supernaturally." The validity of private revelations is open to question, especially if they conflict with empirical observations.

d. Rationalism is the belief that reason yields more reliable knowledge than does revelation or experience. Rationalists emphasize *deductive* reasoning as a primary means of gaining knowledge. Thus, they would argue that mathematics and logic provide more reliable information than does empiricism. Unfortunately, deductive reasoning often begins with a major premise which is either tautologous or so specific that it holds only for a limited number of cases.

2. Several definitions of science were considered: it was described as a source of ultimate authority, a skill, an honorific, or an organized body of knowledge. Science is defined in this book as a method or attitude with the following characteristics:

a. *Reduction and control of bias.* The scientist recognizes his fallibility and makes his observations in as objective a manner as possible. He is not, however, without personal values, principles, prejudices, or interests.

b. *The quest for precision.* A distinction was made between the scientist's desire for precision and accuracy and the quest for absolute or irrefutable knowledge sometimes erroneously attributed to him. The scientist depends upon his tools, techniques, and methods to provide himself with knowledge, but he recognizes that this knowledge is open to question and verification.

c. *Verification.* Because the scientist is fallible, he must allow others to verify or disprove his conclusions. In turn, he examines evidence provided by other investigators. It was pointed out that verification is possible only

a classroom, which may involve highly complex skills, and accumulating knowledge about education, which requires other skills.

5. Science is composed of both descriptions and explanations. Descriptions may be either molecular or molar, depending upon the purposes of the investigation and the amount of information the researcher is willing or able to assimilate. Molecular descriptions are minute, detailed, and specific; molar descriptions are more general and encompass broader aspects of behavior.

Explanations show how an event or phenomenon is related to a general law or principle. One way of explaining phenomena is to relate observations to a body of theory. Another way is to develop a model and show the relevance of the observation to the model. Four explanatory models were described: the deductive model involves drawing specific conclusions from premises, the probability model involves induction and prediction, the teleological model explains events in terms of purposes or goals, and the genetic model explains events by specifying their antecedent conditions.

PRACTICE EXERCISES

1. Each of the following statements might be considered by some persons as "perfectly obvious," and therefore not in need of investigation. Indicate how certain you are that each statement is "correct." If you are uncertain that it is correct, indicate what evidence you would need to determine the correct answer. If you are certain the statement is correct, indicate the reasons for your belief.

a. In a fourth-grade class, John is older than Mary, Mary is older than Peter, and Peter is older than Bill. Of the four children, Bill is the youngest.

b. By lowering the pupil-teacher ratio, teachers will be able to do a better job of teaching than they are now able to do.

c. Kindergarten teachers should have had children of their own to raise before they enter the teaching profession.

d. The teacher should be a model citizen so the children will emulate him.

e. Some children love their teachers.

f. Schools are either elementary, junior high, high schools, or colleges.

2. Indicate the difficulty (if any) in accepting each of the following persons or statements as authoritative. If further information is needed, indicate what type of information you would require before you could evaluate the statement or person.

a. Judge Jones, a member of the Supreme Court, states that the poor relationship between parents and children is the cause of juvenile delinquency.

b. The author of this textbook says that a distinction can be made between teaching as an art and education as a science.

c. Admiral Rickover states that the schools in Europe are better than the schools in the United States.

d. John Dewey says that values depend upon their consequences.

- c. Most educational psychologists agree that motivation is important in learning.
3. For each of the following examples, indicate the methods you could employ to either verify or reject the statement or question. Which statements or questions cannot be investigated? Why?
- How high is up?
 - The principal of a school is responsible for its program.
 - Teachers should be equally prepared in knowledge of subject matter and in methods of teaching.
 - The kindergarten teacher should receive the same pay as the high school teacher if they are both equally well trained to do their jobs.
 - These questions are not difficult.
4. Indicate the underlying assumptions for each of the following statements.
- The curriculum should be based on the Great Books.
 - John has an IQ of 84.
 - Children should be required to work in the cafeteria at school.
 - Sooner or later science will answer all of the pressing educational problems.
 - Reading is more important than playing.
5. What theory could you develop that would help explain each of the following groups of facts? Assume that each of the facts is "true."

Facts for Group I

- At 6:00 A.M., John is awakened by a person talking to him very softly from the window of his room.
- The voice is so low that he cannot make out what the person is saying.
- There is no one in his room but himself.
- He runs outside but there is no one there.
- Each time he returns to the room he hears the voice again.

Facts for Group II

- Of 20 cats isolated from birth from all other animals, 9 were known to kill rodents.
- Of 21 cats who saw rodents being killed, 18 were known to kill.
- Of 18 cats reared with rodents, only 3 were known to kill.
- The cats' diet (vegetarian or nonvegetarian) was unrelated to whether they killed or not.³⁰

Facts for Group III

- Most of the world's best-known persons have been males.
- More men than women have been in homes for the retarded.
- More men than women have been identified as gifted.

Facts for Group IV

- Negroes who migrate to the North do as well on standardized tests as those who remain in the South.
- The average tested IQ for the Negro is 9 to 10 points lower than the average Caucasian's.

³⁰Z. Y. Kuo, "The Genesis of the Cat's Response to the Rat," *Journal of Comparative Psychology*, XI, No. 1 (1930), 1-35.

- c. Some Negro children have IQ's as high as 200.
 - d. Northerners consistently score higher than Southerners of the same racial extraction.
 - e. Caucasians score consistently higher than Negroes living in the same region.
6. For each of the following theoretical positions, indicate what facts you would need to give credence to the theory. How could such facts be obtained?
- a. *Theoretical Position I*: Frustration leads to aggression. (Frustration-Aggression Theory)
 - b. *Theoretical Position II*: For learning to occur there must be a reinforcement. (Theory of Reinforcement)
 - c. *Theoretical Position III*: Much significant human behavior is unconsciously motivated. (Theory of Subconscious Motivation)
 - d. *Theoretical Position IV*: Children are born bad. (Theory of Innate Depravity)
7. Indicate what type of explanation is most called for in each of the following questions (deductive, probability, teleological, or genetic):
- a. Does John smoke because he has an oral fixation?
 - b. Will that door open if I turn the handle to the right and push it?
 - c. Did Columbus land in Santo Domingo in 1492?
 - d. Why were schools built?
 - e. Can safety belts help prevent injury?
 - f. Why does X^2 times X^3 equal X^5 ?
 - g. Why are the streets wet when it rains?

SELECTED RELATED READINGS

1. Feigl, Herbert, and May Brodbeck, *Readings in the Philosophy of Science*. New York: Appleton-Century-Crofts, 1953, 811 pages. A collection of readings ranging widely in level of difficulty. Students will find the following selections most helpful in extending their knowledge of the meaning and assumptions underlying science: Pages 8–18 contain an excellent critique of misconceptions of science; advanced students may find selections on pages 47–102 concerning the meaning of verifiability to be of interest. The meaning of *explanation* in the sciences can be found on pages 319–352 and 688–743; concepts of causality and determinism are treated on pages 387–407. Of special importance to students of educational psychology is Section V, on “Philosophical Problems of Biology and Psychology,” pages 523–659. Pages 757–780 contain some excellent articles on “The Limits of Science,” “Causality and the Science of Human Behavior,” and “The Laws of Science and the Laws of Ethics.”
2. Kaplan, Abraham, *The Conduct of Inquiry*. San Francisco: Chandler Publishing Company, 1964, 428 pages. Written by a philosopher, this text examines the philosophical assumptions underlying research in the behavioral sciences.

3. Larrabee, Harold A., *Reliable Knowledge*. Cambridge, Mass.: The Riverside Press, 1945, 685 pages. An excellent introduction to the problems of obtaining and evaluating reliable knowledge. Especially relevant material can be found in Chapter 2 (Man as Knower: From the Inside); Chapter 3 (Formal Logic—What Follows from Premises); and Chapter 4 (Toward Scientific Method: Observation—"Get the Facts").
4. Nagel, Ernest, *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World, Inc., 1961, 618 pages. Students will find the introduction, titled "Science and Common Sense," to be of interest. In addition, Chapter 2, "Patterns of Scientific Explanation," should extend the present discussion of the role and meaning of *explanation* in the sciences.

CHAPTER TWO

The Role of Research in Education

At the First Annual Phi Delta Kappa Symposium on Educational Research, the late Palmer Johnson observed, "In a population which is so dependent upon research, it is sad to reflect how few people perceive what it is all about. *Research is an approach to a comprehension of the universe along a broad thoroughfare of organized knowledge solidly established on observation and experiment imbedded in a matrix of theory.* It is a highway that is continuously being lengthened, widened, and mended."¹

However, Professor Johnson's definition of research appeared overly idealistic to the editor of the *Phi Delta Kappan*, "a journal for the promotion of research, service, and leadership in education," who stated, "That part of the research highway devoted to education is, it seems to us, in exceedingly poor repair, full of holes attributable to amateur workmanship, and breaking up entirely in spots because of bad engineering and faulty specifications."²

These two quotes emphasize an increasing concern among educators when they discuss educational research. Such discussions are likely to extol the promise of research in general, at the same time admitting that, in education, the promise has not been entirely fulfilled.

This paradox of extolling the promise of research while criticizing its accomplishments is based upon a misunderstanding of the nature of educational research. Although Americans have long emphasized "Progress . . . to be the first law of nature,"³ they have, at the same time, refused to state explicitly the direction that progress ought to take. Inherent in this plea for

¹Palmer O. Johnson, "Introductory Remarks at Opening of the Symposium on Educational Research," in *First Annual Phi Delta Kappa Symposium on Educational Research*, ed. Frank W. Banghart (Bloomington, Ill.: Phi Delta Kappa, Inc., 1960), p. xv. Italics added.

²Stanley Elam, "What—and Where—is Educational Research?" *Phi Delta Kappan*, XLI, No. 6 (1960), 241.

³Vernon L. Parrington, *The Romantic Revolution in America*, Vol. II of *Main Currents in American Thought* (New York: Harcourt, Brace & World, Inc., 1954), p. ix.

progress is the belief that research can provide, or at least should provide, the knowledge needed to contribute to educational reforms and improvements. What has not been understood is that research alone is unable to transform the schools into future Utopias.

Palmer Johnson was correct when he stated, "Research is *an* approach to the comprehension of the universe . . ."; but research is not the *only* approach, and it is not a panacea for all educational problems. Changes in education have resulted from and contributed to many diverse social, economic, and political conditions. Research does not exist in a vacuum unaffected by the hopes, aspirations, and limitations of the rest of society. Considering the fallibility of research findings and remembering that these findings must be evaluated in the light of local economic, political, social, and aesthetic beliefs and practices, it should be clear that any given research finding cannot expect immediate acceptance and implementation in all schools throughout the nation. What may be a reasonable way of dealing with a problem in one community may not be desirable or feasible in another. Thus, although research may provide knowledge about empirical relationships, educators may very well refuse to implement these findings if, for example, the implementation might lead to economic or social injustice.

TYPES OF RESEARCH ACTIVITIES AND THEIR RELATIONSHIP TO EDUCATIONAL CHANGE

A distinction may be made between *basic* research, whose purpose is to develop a body of theoretical knowledge, and *applied* research, which is designed to answer questions having immediate applicability. The desire to produce innovations usually calls for some form of applied research. Later in this chapter we will show that there is a reciprocal relationship between basic and applied research and that they are neither antagonistic nor mutually exclusive. However, basic research has too often been neglected and overlooked by those intent upon improving educational practice. Because theory suggests which facts are relevant and necessary to the understanding of a problem, the researcher working outside of a theoretical framework may have only vague principles to guide him in his selection of variables to investigate. Unless there is a background of theory and empirical research to back up some proposed educational innovation, it may be necessary to evaluate every whim and fad in the public schools to see if they will work.

Unfortunately, this is exactly what has happened in education. Instead of first becoming concerned with the development of a broad body of theoretical knowledge, much educational research simply attempts to compare method *A* with method *B*. It is little wonder that educators have become disillusioned about the practical benefits derivable from research when they

realize that there are an infinite number of methods which ultimately could be compared. To discover that *A* is superior to *B* tells us nothing about how *A* will compare with *C*, *D*, *E*, etc. Theory and empirical research may help specify the relationships between these methods so that their common attributes can be investigated and their differences evaluated.

Many of the recent innovations in science and mathematics curricula were developed without previous empirical investigations to back up the claims made for them by their innovators. Empirical studies are needed to help develop theory as well as to validate rudimentary theoretical positions. Today, we have literally hundreds of specific suggestions for teaching arithmetic, ranging from workbooks to Cuisenaire blocks to the currently popular University of Illinois Committee on School Mathematics (UICSM) curriculum. These techniques are being used in schools without sufficient evidence to suggest under what conditions and for what types of students they can be used most effectively. As one researcher put it:

Ideas spring from the innovator's mind into direct application to children and teachers with no intermediate steps. If medicine did things the way we do, the Salk vaccine would have been administered to everyone merely because Jonas Salk thought it should be, and was persuasive enough to get us to do it. . . . We should adopt the point of view that ideas should not be tried on children until there is some large amount of evidence that the ideas are workable and the results will be beneficial.⁴

Behind the statement that innovators should be concerned with research is the principle that one should not try to implement what is not known. Even in the marketing of a new detergent, the way chemicals react on materials must be known before the product is sold to the public. What we have in education is a lack of knowledge coupled with thousands of suggestions on how to do a better job. We need to know what to expect of any innovation before it can be recommended.

Although basic research is necessary, by itself it is not sufficient to justify the use of any proposed innovation. Field tryouts are needed to evaluate the practical effects of the innovation. It may turn out, for example, that a new teaching approach has much to recommend it theoretically but is so expensive to implement that it would be unworkable under present conditions. The responsibility of *development* and *demonstration* is to make proposed changes practicable.

Development as an impetus to innovation

Development includes those activities used to lower costs and increase the efficiency of products, practices, or conditions in order to promote their greater usage. In industry, for example, the responsibility of development is to take a product which has been shown to have value and produce it effi-

⁴Daniel E. Griffiths, *Research in Educational Administration: An Appraisal and A Plan* (New York: Bureau of Publications, Teachers College, Columbia University, 1959), p. 31.

Demonstration as an Impetus to Innovation

Once a product or idea has gone through the development phase of production, it is necessary to *demonstrate* how it works. The purpose of the demonstration is to persuade others to accept the innovation. Development may indicate that a new approach to teaching mathematics is both efficient and inexpensive, but it is up to the demonstration to show teachers and administrators that the approach is operationally useful and desirable; unless it has some appeal which can be demonstrated, it is unlikely to find its way into very many classrooms. As the innovation is demonstrated, new problems for research and development may be generated.

TYPES OF RESEARCH AND THEIR RELATIONSHIP TO INNOVATION

Development and demonstration depend upon having reliable knowledge concerning a proposed innovation. This knowledge can be derived from the application of various research methods: analytic, descriptive, and experimental (see Table 1).

Analytic Research

A number of related research activities may be subsumed under the heading of *analytic research*. This includes mathematical, linguistic, historical, and philosophical analyses, as well as any primarily deductive system that can be used to derive relationships not necessarily of an empirical nature. Examples of analytic research include the analysis of data through statistical procedures, the analysis of language through grammar and linguistics, and the analysis of assumptions and implications through logical and philosophical procedures.

Analytic research helps point out the assumptions and possible consequences of proposed innovations. The historian, for example, may be in a position to show how a proposed change is related to some historical event that has already been tried and evaluated. Analysis may also be used to help establish criteria for the evaluation of an innovation.

One form of analytic research is *data retrieval*. Its purpose is to gather information from both primary and secondary sources. A primary source is one which has direct access to an original observation. For example, an origi-

nal manuscript or a report from an eyewitness are primary sources of information. In contrast, a secondary source is one or more times removed from the original observation. A report from a second-hand witness or a book which summarizes the contents of original reports are secondary sources.

By searching the literature, the investigator concerned with innovation can take advantage of the experiences of others engaged in related pursuits.

TABLE 1

TYPES OF RESEARCH AND THEIR RELATIONSHIP TO EDUCATIONAL INNOVATION

	<i>Analytic</i>	<i>Descriptive</i>	<i>Experimental</i>
Purpose	To derive relationships within a deductive system	To describe existing conditions	To test causal relationships
Methods	Deductive, mathematical, historical, philosophical, legal, linguistic	Correlations, surveys, case studies, direct observation, cross cultural, growth studies	Comparison of experimental and non-experimental groups by systematically varying conditions
Relation to Innovation	Points out assumptions and possible consequences of proposed changes; useful in establishing criteria	Describes currently existing conditions so that they can be modified later	Shows the effects of a proposed innovation

Data retrieval allows the investigator to analyze the relationship between his studies and those of other researchers. It may point out weaknesses in his approach, or help him find a body of related studies to form the basis of relevant theory, thus leading him to select variables for study which can be justified on the basis of prior evidence.

Descriptive Research

The purpose of *descriptive research* is to show conditions as they exist without being influenced by the investigator. Descriptive research encompasses a number of different techniques, including correlational analyses, case studies, surveys, and interviews, as well as direct observation. These techniques will be discussed in greater detail in later chapters.

Descriptive research is often of greatest value during the initial stages of an investigation. If the current status of a problem is described, steps may be taken later to remedy it. If, for example, educators are concerned about the effects that high school dropouts are likely to have on the economy, a first step might be a study to determine how many students leave high school before graduating, what occupations they move into, their salaries, etc. Descriptive research may also help point out the extent of a problem and indicate how serious and widespread it is. Once this information is available, the scientist can begin to resolve the problem.

Experimental Research

The purpose of *experimental research* is to study "causal" relationships. It may be used, for example, to study the effects rewards have on learning or to evaluate the advantages of one method of teaching over another. The investigator tries to control (i.e., eliminate or hold constant) the effects of extraneous conditions in order to study interactions between variables.

Because the experiment studies causal relationships, the effect of any proposed innovation in education is most appropriately evaluated by using some form of experimental design (see Chapter Twelve).

The role of the experiment in solving the school dropout problem might be to determine which of various proposals to retain students leads to the largest number graduated. These proposals might include the establishment of a work-study combination program where the student works half day and attends school half day; or they might involve an evaluation of the "job corps," the use of volunteers to help motivate students planning to drop out of school, etc. Before deciding which proposals will be evaluated, answers are needed to such questions as: "What is the relationship between social class and opportunity for success in the public schools?" "How does the reward structure in schools affect achievement?" "Under what conditions and at what ages do negative attitudes toward school begin?" "What factors are most responsible for motivating students to achieve and enjoy school?" Unless these and other related questions are investigated, an experimental program designed to test one method of reducing dropouts against another is not likely to be very successful.

THE HISTORICAL DEVELOPMENT OF EDUCATIONAL RESEARCH

One reason educators have emphasized the values of research is that it has contributed information needed to implement and evaluate educational practices. The educator, like other professionals, depends upon having sufficient information about his work to act responsibly; it is in the process of obtaining and disseminating this knowledge that research is created.

Prior to the twentieth century, which heralded the beginning of scientific and experimental education, information concerning the curriculum, teacher training, and student growth was based upon the prevailing philosophical beliefs of the time as well as upon the biases of individual educational practitioners. These beliefs went unchallenged until the effects of the technological revolution were felt in the public schools. Technological advances forced the twentieth-century educator to find new ways to cope with the millions of children brought into the public schools by compulsory education laws. He not only had to contend with a greater number of students, but he had to find ways to modify the curriculum so that it could accommodate the ever-increasing list of required subjects forced into it by technological advances.

The scientific movement in education developed out of the need to justify and improve upon existing educational beliefs and practices, and not from a desire to develop a new social science which might parenthetically contribute to educational change.⁵ The educator of the early 1900's (like his modern counterpart) was too busy with practical matters to devote much of his time or attention to founding a science of education, unless that science could have immediate utility. This need for immediate "solutions" led to literally thousands of research activities, with few of them contributing to or derived from theory. Thus, although today we have a great many research studies in education, to a large extent they are not organized into theoretical systems which account for and explain empirical relationships.

The Early Beginnings of Educational Research

Although it is most difficult to determine exactly when the scientific movement in education began, the writings of Johann Friedrich Herbart (1776–1841) may be taken as the first major impetus to the scientific study of education. At Königsberg, where he was a professor of philosophy, Herbart offered a seminar in education and suggested that the scientific study of education was possible. However, his contributions to education were not widely acclaimed until 1895, when the National Herbart Society was founded in the United States. Later, this organization changed its name and became the National Society for the Scientific Study of Education. In 1910 it became known as the National Society for the Study of Education (NSSE), which today provides scholarly yearbooks on a variety of subjects related to education.

Another impetus to the scientific study of education began with the establishment by Wilhelm Max Wundt at Leipzig in 1879 of the first psychological laboratory. Although Herbart had refused to believe that *experimentation* was possible in education (he did claim that *psychology* was both empirical and mathematical), Wundt established laboratory experimentation as a major research tool.

Wundt's experimental laboratory had its American counterparts in William James's laboratory at Harvard in 1875 and in G. Stanley Hall's laboratory at Johns Hopkins University in 1883. James's *Talks to Teachers*, published in 1899, related principles of psychology to problems faced by educators. Hall has been credited with founding of the child study movement in America. In an attempt to discover the "content of children's minds," he popularized the use of the questionnaire, a technique developed by Sir Francis Galton in 1883. By 1891, Hall was able to found *Pedagogical Seminary*. The establishment of this journal was evidence that educators had begun to feel the need to communicate with each other concerning the attainment of knowledge in child growth and development.

⁵H. G. Good, *A History of American Education* (New York: The Macmillan Company, Publishers, 2nd Ed., 1962), p. 415.

The psychology of the early 1900's strongly felt the influence of Darwinism. In 1859, Darwin wrote the *Origin of Species*, which showed that man could be studied by naturalistic observation and that he obeyed the same laws as did other members of the animal kingdom. Thirteen years later, in 1872, Darwin published *Expression of the Emotions in Man and Animals*, in which he pointed out that facial expressions were survivals of various habits related to the continuance of the species. He also suggested that the overt display of an emotion (such as love) could be used to hide its emotional opposite, hate. By 1915, Freud made use of this concept to show how *reaction formations* could be used as a subconscious ego-protective mechanism.

Sir Francis Galton, Charles Darwin's cousin, was greatly influenced by him. Galton was interested primarily in eugenics and the study of individual differences. His studies led him to develop a number of rather crude and simple reaction-time and association tests. In 1882, he was given space in the South Kensington Museum to test, for a small fee, those persons interested in having their abilities measured. He thus established the first testing center, and became, in effect, the first psychometrist. Six years later, in 1888, he published a mathematical procedure for describing the degree of correlation between two variables.

The scientific movement in education was furthered by James McKeen Cattell, who had worked with Wundt at Leipzig, G. Stanley Hall at Johns Hopkins, and Galton in London. In 1890, Cattell used the term *mental tests* in print for the first time, and five years later he was able to persuade the American Psychological Association to evaluate the use of tests in education. Cattell also developed a psychological laboratory at Columbia which encouraged additional investigations into individual differences.

Edward Lee Thorndike, often credited with being the "father of educational psychology," was Cattell's student at Columbia and James's at Harvard. In 1898, Thorndike published *Animal Intelligence*, which showed how animals could be used in a laboratory situation to study intelligence as well as the learning process. The publication in 1904 of his *Theory of Mental and Social Measurements* made him a respected leader both in learning theory and in educational and psychological measurements.

The *measurement* movement provided fertile ground for two types of research. First, there were a number of immediate and practical needs requiring the use of tests. In France, for example, compulsory education laws forced many children who had not attended school before to enter the public schools. Many of these children were so retarded intellectually that ways had to be found to separate them for special classes. Binet and Simon undertook the task of developing a test that could be used for differentiating pupils into regular and special classes. Their first practical scale was devised in 1905. With the development of this scale, many other practical uses for tests were found. By World War I, mental measurement experts were well accepted in both schools and industry.

Second, tests could be used to help develop *theories* about the nature of intelligence or personality. In 1902, for example, Thorndike suggested a *multifactor theory of intelligence*. Intelligence, he argued, was composed of linguistic, mechanical, and social factors or abilities. Two years later, Spearman devised a method called *factor analysis* for analyzing complex batteries of tests; he hypothesized a *two-factor theory of intelligence*. According to this theory, intelligence could be described by a *g* (general) factor and *s* (specific) factors. The Thorndike-Spearman controversy was important because it pointed out the need for systematic theories of intelligence and for more sophisticated ways of analyzing tests. In 1938, Thurstone showed that intelligence could be analyzed into a number of *primary mental abilities*; he thus supported Thorndike's earlier contention that intelligence was not a unitary trait.

Knowledge about the learning process began with the studies of Hermann Ebbinghaus in 1885. By memorizing lists of nonsense syllables (bej, fup, hif), he plotted his own learning curves and compared them with curves obtained by memorizing meaningful information. From these comparisons and from other investigations, he showed the advantages of distributed over massed practice and of the whole method of learning over the part method. The Bryan and Harter experiment on the acquisition of skills needed to learn telegraphy (1897) furthered Ebbinghaus' work. Bryan and Harter were able to describe "typical" learning curves and to suggest various reasons for *plateaus*, or trials where little progress was made.

However important the Ebbinghaus and Bryan and Harter investigations were, Thorndike's *Connectionism* is generally credited with being the first major theoretical position on learning. Thorndike provided evidence that learning could be facilitated by the application of various "laws," such as the law of effect (learning is facilitated if it is followed by a state of satisfaction and hindered if followed by a state of dissatisfaction). His experimental work encouraged his students and colleagues to develop their own modifications and improvements on his original system. Pavlov, for example, knew about Thorndike's studies of animal learning at the time he was working on conditioned reflexes and gave credit to him for his early research on animal learning.

The Rise of the Psychological "Schools"

Wundt's experimental laboratory at Leipzig attempted to study the contents of the mind, through introspection and experimentation. This was a reaction against the type of armchair philosophy which had become associated with psychology in its early development. The new psychology, called *structuralism*, accepted the following premises: 1) psychology can be investigated scientifically; 2) psychology is defined as the study of mental contents; 3) mental contents can best be investigated through introspection or the careful analysis of one's own sensations and perceptions in response to specific stimuli; and 4) complex experiences can be analyzed into simpler ones.

Structuralism found its way to the United States with one of Wundt's students, Edward Titchener, who directed the psychological laboratory at Cornell until his death in 1927. Agreeing with Wundt on the purpose and methods to be used by psychologists, he rebelled against the then-current school of *functionalism* led by John Dewey and James Rowland Angell. Instead of describing the content of the mind, functionalists studied its functions. That is, they were more interested in the purpose and uses of intellect than in ways intellect might be structured.

The battle between structuralism and functionalism had important repercussions for education. If mind were structured, one responsibility of the educator would be to discover what kind of mental structure children have. In contrast, the functionalist was more concerned with purpose, or means-end relationships. He did not ask what structures the mind was composed of, but how intellect functions and what it might be expected to do.

Although Dewey was trained as a philosopher, he had become influenced by G. Stanley Hall, and was elected president of the American Psychological Association in 1897. In 1902 he became the director of the School of Education at the University of Chicago; in 1904 he went to Teachers College, where he remained as professor of philosophy until he retired in 1931.

Dewey's 1896 paper, "The Reflex Arc Concept in Psychology," showed his dissatisfaction with structuralism. In this paper, he gave an example of a child who sees a flame, reaches for it, and gets burned. Dewey argued that it was necessary to consider the entire act as a continuous one, and that the separate elements of perception, movement, and sensation have no meaning by themselves. One cannot separate the elements of an act from its consequences or purpose. Essentially, this meant that Dewey did not consider valid many of the dualisms psychology had inherited from Descartes. Thus, he eliminated such dichotomies as those between mind and body, method and content, or ends and means. These conceptions eventually led him to write *Democracy and Education* in 1916 and *Experience and Education* in 1938. These texts provided the basic tenets of progressive education.

James Rowland Angell was brought to the University of Chicago by Dewey in 1894 and became its president 24 years later. In 1906 he was elected president of the American Psychological Association; his presidential address was titled "The Province of Functional Psychology." At Chicago, Angell developed a strong department of psychology dedicated to the principal tenets of functionalism.

A third psychological school was *Behaviorism*, which developed in opposition to structuralism and functionalism. It opposed structuralism on the grounds that psychology should not concern itself with consciousness or introspection; it opposed functionalism by stating that the proper subject matter of psychology was the study of behavior, not mental functions. While structuralists were not opposed to the study of behavior, they argued that this subject was biology and not psychology.

Certainly the best known and most influential of the early behaviorists was

John Watson, who received his doctorate in psychology from the University of Chicago in 1903, just one year before Dewey left Chicago for Teachers College. Watson was interested in animal studies and could not see the application of introspection or the analysis of consciousness to his work. Since both structuralism and functionalism depended at least in part on these methods, Watson became increasingly dissatisfied with both, and in 1913 he formally founded a new school by publishing a paper titled "Psychology as the Behaviorist Views It." The purpose of all sciences, he argued, was prediction and control, and the appropriate subject matter of psychology was behavior. The study of consciousness was unscientific and therefore was not to be considered a part of psychology.

Watsonian behaviorism had a number of important implications for education. If, for example, the purpose of psychology is to control behavior, teachers might be able to develop ways of controlling the behavior of children in the classroom. To accomplish this, the child's environment would have to be structured and to be controlled, not only in school, but at home as well. Learning could be controlled by means of the conditioned response; the teacher would have control over the stimulus. Because of Watson's insistence that psychology could be useful, some applied areas of psychology received an important impetus.

Watson denied the concept of "innate intelligence." Intelligence, he believed, was acquired in early childhood. This belief led him to study the development of emotions and how they might be modified through conditioning. Only three emotions could be found in infants—rage, fear, and love. Rage could be produced by restricting bodily movements, fear was induced by loud noises or loss of support, and love was demonstrated in the smiling and cooing response to tactile skin stimulation. Watson was so certain that heredity played essentially no role in determining behavior that, in 1925, he made his famous challenge:

Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I'll guarantee to take any one at random and train him to become any type of specialist I might select—doctor, lawyer, artist, merchant-chief and, yes, even beggar-man and thief, regardless of his talents, penchants, tendencies, abilities, vocations, and race of his ancestors.⁶

Although his challenge was never accepted, educators saw in it the unlimited possibility of training pupils for whatever objectives they had in mind. One of the most practical consequences of this point of view was the development by Pressey in 1924 of the teaching machine, which presented stimuli in the form of carefully constructed lessons or programs that would reinforce the learner if he responded correctly to the questions. Later, in 1953, Skinner published *Science and Human Behavior*, in which he showed not only how the

⁶John B. Watson, *Behaviorism* (New York: W. W. Norton & Company, Inc., Publishers, 1926), p. 82. Copyright 1925 by The People's Institute Publishing Company, Inc., and 1953 by John B. Watson.

away from an excessive dependence on the role of memory and drill, educators eagerly embraced the Gestalt movement as a panacea for many of their problems. However, most American psychologists considered themselves to be behavioristically oriented, and they dominated the research scene in the behavioral sciences. The result was that relatively little research on learning was done using Gestalt principles. In recent years, the *cognitive theorists* have done much to bridge the gap that existed for many years between behaviorists and Gestaltists.

The *psychoanalytic* school owes its origins to Sigmund Freud. Although G. Stanley Hall is usually credited with being the "father of child psychology," it was Freud who provided the most comprehensive theory of child behavior. Psychoanalytic theory developed out of Freud's experiences in working with neurotics. According to Freud, all behavior is motivated, sometimes consciously and sometimes subconsciously. Man, he said, is not always aware of his reasons for acting as he does; he may repress the reasons for his behavior. Repression was considered to be an ego-protective mechanism, along with sublimation, identification, compensation, etc.

Freud argued that the purpose of education was to teach the child to constrain his "natural instincts." To do this, the school has to provide ways for the child to *sublimate* his aggressive and destructive tendencies. It cannot simply constrain or control him, because this can lead to neurosis. Freud believed that the "ideal" form of education could teach the child to curb his own antisocial tendencies without creating neuroses.

The psychoanalytic movement depended heavily on case studies or comprehensive descriptions of the individual's past experiences to explain present behavior. Behaviorists, in general, could agree with this point of view easily by arguing that the totality of one's present behavior is a combination of his earlier experiences. But to the Gestaltists, this was too much like arguing that a person is simply the sum of his individual experiences, since they believed that the whole was different from the parts.

The Dissolution of the Psychological Schools

The psychological schools arose to challenge the prevailing orthodoxy. The functionalists organized as a protest against the structuralists and in turn were attacked by behaviorists and Gestaltists. As each school developed, it created its own research methods to fit in with its philosophic orientation. Thus, introspective methods were considered absolutely necessary by Titchener; for Watson, only the objective description of stimulus-response relationships would do. Case history methods were of primary importance to the psychoanalysts, but were opposed by the Gestaltists. By developing its own methods and techniques, each school made its contribution to the studies of education and psychology.

The psychological schools were largely the products of a few individuals: Wundt and Titchener for the structuralists; Dewey and Angell for the func-

Nonetheless, to some extent, one can find traces of all of these psychological schools still present in the writings of modern psychologists and educators. While practically no one is willing to call himself an introspectionist any more (it is considered to be too "mentalistic"), verbal descriptions of experiences still play a most important role in modern research. Questionnaires, for example, depend upon having subjects respond to questions about their beliefs or attitudes, and thus are a form of introspection. The functionalists stressed the use of tests to measure the purposes and uses of intellect, and the psychoanalysts used the case history methods. Each of these techniques will be described in greater detail throughout the remainder of this text.

Research in Educational Administration

The problems of organizing, financing, and operating public schools are complex. Before the twentieth century, empirical research in educational administration was almost nonexistent, since the impact of scientific method, especially in the development of psychological tests, had not yet been felt. Nineteenth-century administrators developed their methods by depending upon *a priori* value judgments and what was considered to be common sense. Wundt's admonition that *a priori ist garnichts* (unempirical claims are meaningless) had little effect on the development of educational administration.

The motivation to investigate administrative problems empirically came from two sources: industrialization and the need for reform.⁷ Industrializa-

⁷Raymond E. Callahan and H. Warren Button, "Historical Change in the Role of Man in the Organization: 1865-1950," *Behavioral Science and Educational Administration*, 63rd NSSE Yearbook, Part 2 (Chicago: University of Chicago Press, 1964), pp. 73-92.

tion and large-scale business operations had penetrated into almost every aspect of American economic life by the early 1900's and had strongly influenced American values and practices in general. Along with these changes in values came the demand for greater reform and efficiency in public school operations. Unfortunately, school administrators were not in a position to cope with many of the practical problems brought about by compulsory education laws, increased enrollments, and the demand for economy. If businesses could operate efficiently and economically, the public could see no reason for school administrators to do less well. Faced with the necessity of increasing school services while reducing costs, administrators looked to research to resolve their problems. What they meant by research, however, did not include theory. If there was to be any research, it would have to be judged by the extent to which it could reduce public criticism and increase school efficiency. Theory might be tolerated, but not demanded.

The school district reorganization movement is a case in point. The idea of consolidating small, inefficient, and relatively expensive school districts into larger, more economical ones was suggested as early as 1837 by Horace Mann, but empirical research methods didn't find their way into educational administration until 1910. In that year, Calvin N. Kendall, the Superintendent of Schools in Indianapolis, Indiana, surveyed the Boise, Idaho, school system. The school survey is a type of descriptive research which utilizes statistics on school trends in enrollments, salaries, space utilization, curriculum, etc., in order to implement recommendations leading to more efficiently operated schools. With the development of the school survey, it was possible to relate size of organizational unit with cost per pupil instructed. Publications by the United States Office of Education⁸ and the National Education Association⁹ presented data necessary to understand how school districts are organized and why small schools are unable to have a wide variety of programs and curricular offerings.¹⁰

Research was used to understand the relationship between reorganization and finance problems involving political and legal statutes, and the sociological effects that reorganization would have on the local community.¹¹ It was also used to evaluate the effects of school district reorganization. Studies were conducted to evaluate the effects of district consolidation on educational opportunities, academic achievement, social and personal behavior of children, the financial structure of states and districts, and the community itself.¹²

⁸Walter H. Gaumnitz, *Small Schools are Growing Larger*, U. S. Department of Health, Education, and Welfare, Office of Education, Circular No. 601 (Washington, D. C.: Superintendent of Documents, U.S. Government Printing Office, Sept. 1959).

⁹M. C. S. Noble and H. A. Dawson, *A Handbook on Rural Education* (Washington, D. C.: Department of Rural Education, National Education Association, 1961).

¹⁰Research Division of the National Education Association, "Subjects in Small High Schools," *NEA Research Bulletin*, XL, No. 2 (1962), 56-58.

¹¹Burton W. Kreitlow, "Organizational Patterns: Local School Districts," *Review of Educational Research*, XXXI, No. 4 (1961), 382-384.

¹²Kreitlow, "Organizational Patterns," pp. 384-388.

Although most of these studies did provide evidence in favor of the reorganization movement, they did not contribute substantially to a theory of educational administration. Thus, while a great deal of effort was expended in resolving an immediate problem, it was expended primarily to justify a proposed reform and not to contribute to an understanding of administrative organization through basic research. Griffith, himself an administrator, has been particularly critical of this type of administrative research:

Without doubt, the greatest weakness of research in educational administration is the lack of theory. Most studies in educational administration are done at the level of "naked empiricism." By this we mean that the researcher has an idea that a vaguely defined problem needs to be solved. He collects data through a questionnaire, survey, or some other method and attempts to find an answer by "looking at the data." By following this procedure we have amassed tons of data, but have come up with very few answers.¹³

It was the period following the close of World War II that marked the inception of theory into educational administration. Prior to that time, what little administrative theory existed was largely the work of political scientists, management executives, engineers, and industrial researchers. But in 1946 and 1947, there were three events in education that helped clear the way for theoretical research.¹⁴ First, the Kellogg Foundation's education advisory committee recommended that the Foundation sponsor research into the study and improvement of educational administration. Second, the American Association of School Administrators (AASA) included the professionalization of the school superintendency in its objectives for the Association. And third, professors of educational administration founded the National Conference of Professors of Educational Administration (NCPEA).

In 1950, the Kellogg Foundation instituted its Cooperative Programs in Educational Administration (CPEA) in five major universities. A year later three more universities were added. These centers helped attract scholars from other disciplines to study educational administration and encouraged capable graduate students to enter the program. While these centers did not convert educational administration into scientific laboratories, they did encourage a scientific approach to the study of administrative problems. In 1955, the Kellogg Foundation supported the University Council for Educational Administration. This organization consisted of the original eight universities involved in the CPEA, plus other universities offering doctoral work in school administration. Its purpose was to conduct cooperative research on administrative problems, as well as to conduct scholarly seminars, publish case studies, and disseminate papers on important subjects.

In addition to the Kellogg Foundation, the AASA has exerted a powerful

¹³Griffiths, *Research in Educational Administration*, p. 16.

¹⁴Hollis A. Moore, "The Ferment in School Administration," *Behavioral Science and Educational Administration*, 63rd NSSE Yearbook, Part 2 (Chicago: University of Chicago Press, 1964) p. 15.

influence on the training of administrative personnel. This group, composed of superintendents and assistant superintendents, was discontented with college-level training programs in administration. In 1947, it requested \$75,000 from the Kellogg Foundation to study the role and training of school superintendents. Although this proposal was rejected by the Foundation, it agreed to the development of the CPEA.

The third major contribution to the change of character in administration was the formation of the NCPEA. This organization developed in 1947 out of a meeting of the AASA, and provided a means whereby college professors in administration could voice the need for basic research. In 1957 it published *Administrative Behavior in Education*, edited by Roald F. Campbell and Russell T. Gregg, which pointed clearly to the scientific road which educational administration could take. The Midwest Administration Center at the University of Chicago has also taken a leading role in developing a science of administration.

Research on Teaching Methods and the Curriculum

Empirical research involving analysis and evaluation of the elementary and secondary curriculum, like research on educational administration, was practically nonexistent before the twentieth century. Only after suitable research tools, techniques, and methods had evolved was it possible to scrutinize the assumptions that were generally accepted by nineteenth-century schoolmen. These assumptions concerned the nature of the learner, the characteristics of learning, and the objectives which were to be learned.

The teaching methods of the nineteenth century emphasized the role of memorization in the learning process. Combined with the concept of formal discipline, the memoriter system demanded that all students learn difficult subjects in order to "train the mind like a muscle." The result was a sterile, frustrating, and ineffective curriculum for a majority of the pupils.

The motivation to improve learning conditions did not originate with the researchers but with the general public. As long as the schools emphasized the value of memorization and insisted upon teaching "hard" subjects to "exercise the mind," teachers found it necessary to fail large numbers of students who were either unwilling or unable to profit by the instruction. This state of affairs was intolerable to the rising working class, which realized the necessity of a free and universal educational system to their own social and economic conditions. The memoriter system could not fit into this pattern.

The demand for a free and universal public educational system presumed a curriculum which could be adapted to the abilities and interests of the majority of school-age children. Much of what the public schools taught was considered to be rather useless for these children. In place of the memoriter system, educators proposed using the criterion of "social usage."¹⁵ The

¹⁵H. G. Good, *A History of American Education* (New York: The Macmillan Company, Publishers, 2nd Ed., 1962) p. 421.

As in spelling so in arithmetic, scientific research grew out of the practical problems of life and education. It was an effort to test unverified beliefs and to make vague and general statements precise and specific. It is not correct to say that educational reform was begun by the scientific investigators. On the contrary, they continued the trend already set in motion.¹⁷

Research was also needed to evaluate the claims made by those who favored the memoriter system and the principles encompassed by the concept of formal discipline. As early as 1890, William James had experimentally investigated formal discipline and was able to state, "... when schoolboys improve by practice in the case of learning by heart, the improvement will, I am sure, be always found to reside in the *mode of study of the particular piece* . . . and not at all to any enhancement of the brute retentive power."¹⁸ In 1901, Thorndike and Woodworth published their important paper on transfer of training, which did much to destroy the presumed advantages claimed for formal discipline.¹⁹

As in educational administration, the first investigations into the curriculum and teaching methods were surveys of current practices. In 1907, for

¹⁶W. W. Charters, *Curriculum Construction* (New York: The Macmillan Company, Publishers, 1923).

¹⁷Good, *A History of American Education*, p. 415.

¹⁸William James, *The Principles of Psychology* (New York: Holt, Rinehart & Winston, Inc., 1890) pp. 664-665.

¹⁹E. L. Thorndike and R. S. Woodworth, "The Influence of Improvement in one Mental Function Upon the Efficiency of Other Functions," *Psychological Review*, VIII, No. 3 (1901), 247-261.

example, Thorndike²⁰ showed that almost 82 per cent of the students who had entered school between 1900 and 1904 were eliminated from these schools by the ninth grade. Two years later, Ayres²¹ reported that only 10 per cent of the students who began school graduated from high school. These investigations implied that an immediate reform in all aspects of the curriculum was needed.

The movement for educational reform had two general effects. In the first place it encouraged the development of new innovations and practices to bring about needed changes. The literature of the 1920's and 1930's was replete with new ideas, ranging from the development of core programs to interrelate subject fields, to the initiation of plans whereby individual students would "contract" for assignments. In the same era can be found innovations in the curriculum of the public schools. The combined effects of these changes did much to reduce dropout rates and increase learning efficiency, but did not add much to a science of education.

The second effect of the reform movement was to encourage research into all areas of the curriculum. The motive was to justify the use of an innovation and to demonstrate how efficient or useful it might be. Literally thousands of studies were published which compared this teaching method against that or which showed the advantages of one curricular innovation against another. Very few studies were published which had theory to guide them, either in selecting variables to compare or in determining what new investigations were needed. Instead, much of educational research was trivial, repetitive, or poorly designed.

Unfortunately, these conditions have not changed markedly even today. In 1964, for example, Ohles wrote, "... what passes for research today is a frenzied effort to prove (not test) an innovation."²² Since there is literally no end to the number of new ideas one can dream up, it follows that demonstrations and development projects alone are not likely to add to basic knowledge on how children learn. That few of these innovations have gained widespread support or acceptance attests to the weakness of the approach.

In recent years, a new program has permeated the curriculum. In 1951, Beberman and others at the University of Illinois began working on the secondary mathematics curriculum. This project was called the University of Illinois Committee on School Mathematics (UICSM), but is more commonly referred to as the "new mathematics." It was designed to make mathematics more comprehensible, less laden with unexplained rules, and of greater

²⁰E. L. Thorndike, *Elimination of Pupils From Schools*, U. S. Bureau of Education Bulletin (Washington, D. C.: U. S. Government Printing Office, 1907).

²¹Leonard D. Ayres, *Laggard in Our Schools* (New York: Russell Sage Foundation, 1909).

²²John F. Ohles, "The Futility of Change?" *The Educational Forum*, 29, No. 1 (1964), p 16. By permission of Kappa Delta Pi, an Honor Society in Education, owners of the copyright.

interest to students. Seven years later, the National Science Foundation supported the School Mathematics Study Group (SMSG), which prepared its own materials for use in the public schools. Similar projects are currently being developed by a number of other groups. Harold Hand has had this to say about these projects:

Without exception all of the news items in the press and all the comments in the popular magazines concerning the new mathematics programs that I have encountered have been laudatory to some degree. . . . About three years ago, however, the distinguished head of the mathematics department in a large eastern university made a blistering attack on the new mathematics courses. . . . His attack was largely based on the grounds that the new courses neglected aspects of mathematics which youngsters need to learn. This was followed by a three page statement, also critical of the new mathematics programs . . . which was signed by professors of mathematics at 32 institutions.

Neither school board members nor the public at large has been told that there are competent scholars in mathematics who seriously question the soundness of the highly touted mathematics courses. Nor have they been told that the experimental validation of a belief on which the new programs in science as well as those in mathematics are banking is lacking.²³

It would be unfair to characterize all curricular and instructional research as being without theoretical justification, although it is true of the great majority. Fortunately, recent years have shown an increasing concern by educators to develop theoretical models and to test these models empirically. In 1962, for example, Gagné²⁴ prepared a theoretical model to help in understanding the process of solving mathematical equations based upon learning sets. Piaget's work on problem solving,²⁵ Guilford's analysis of intellect,²⁶ Bruner's work on instructional theory,²⁷ Carroll's analysis of language instruction,²⁸ Suppes's study of mathematical reasoning,²⁹ and Suchman's work on inquiry training in science³⁰ are just a few of the recent contributions to educational theory.

²³Harold C. Hand, "Integrity and Instructional Innovation," *The Educational Forum*, 30, No.1 (1965), 9-10. By permission of Kappa Delta Pi, an Honor Society in Education, owners of the copyright.

²⁴Robert M. Gagné, "The Acquisition of Knowledge," *Psychological Review*, LXIX (July 1962), 355-365.

²⁵Jean Piaget, *Logic and Psychology* (New York: Basic Books, Inc., Publishers, 1957).

²⁶J. P. Guilford, "The Structure of Intellect," *Psychological Bulletin*, 53, No. 4 (1956), 267-293.

²⁷Jerome S. Bruner, *The Process of Education* (Cambridge, Mass.: Harvard University Press, 1960).

²⁸John B. Carroll, "The Prediction of Success in Intensive Foreign Language Training," in *Training Research and Education*, ed. Robert Glaser (Pittsburgh, Pa.: University Press, 1962).

²⁹Patrick Suppes, "Toward a Behavioral Psychology of Mathematics Thinking," Unpublished manuscript, Stanford University, 1963.

³⁰J. Richard Suchman, "Inquiry Training: Building Skills for Autonomous Discovery," *Merrill Palmer Quarterly*, VII (1961), 147-169.

FACTORS HINDERING EDUCATIONAL RESEARCH

Although research has been able to make valuable contributions to knowledge about education, it has nevertheless been unable to make the type of contribution which many educators have felt it should. As we stated earlier, educators have extolled the promises of research but have been disappointed in its results. In part, this is due to the fact that much of educational research has been used to evaluate innovations without developing basic or theoretical knowledge about education in general. There are, however, other conditions that have hindered the progress of educational research. These include the lack of funds for research, the general lack of skill among educators in the methodology of research, the difficulties in bridging the gap between theory and practice in the schools, and the willingness of some educators to accept unreliable or undependable research findings.

The Lack of Funds for Educational Research

Educational research cannot make the progress demanded of it until adequate funds are made available for equipment, manpower, and supplies. To be sure, even if sufficient funds were provided, research would still be hampered by inadequately trained researchers and the use of inefficient research designs. Nonetheless, without adequate expenditures for research, the best designs or the best researchers are likely to fail.

Industry and business have been estimated to place between 5 and 15 per cent of their funds into research. By carefully explaining the values of research to stockholders and governing boards, industries have been able to appropriate rather large sums of money for both basic and applied problems. Unfortunately, it has not been as easy for educational administrators to urge school boards into expending larger proportions of their budgets for research.

Fortunately, however, a number of agencies have been willing to provide funds for educational research. In 1954, for example, the federal government was authorized under the Cooperative Research Program of the Office of Education to spend almost \$1 million to enter into contracts with universities and state departments of education. These amounts were raised in subsequent years, so that by 1964 the federal government had appropriated approximately \$34 million under the Cooperative Research Program; by 1966 some \$70 million had been allocated by Congress.

These appropriations, though seemingly large, are still rather small; it has been estimated that less than 0.05 per cent of the total annual educational budget has been granted for educational research. Compared with this, industry, business, and private foundations have been rather generous in providing research funds for education. In 1962, the Ford Foundation was

estimated to be granting more money to improve education than was the federal government.³¹

The federal government alone cannot provide the full support needed if research is going to live up to the demands placed upon it by educators. Other agencies (both public and private) should share the responsibility for providing adequate research funds. For example, universities and colleges are estimated to provide anywhere from one-fifth to one-third of their budgets for research.³² Public schools, in contrast, ordinarily devote 2 per cent of their budgets, and often less, to these activities.

Lack of Training in Educational Research

One important reason that educational research has not made the contribution educators have demanded lies in the training of researchers themselves. Competent researchers simply do not develop in a vacuum—they must be selected with care and allowed to work and study in an atmosphere which respects intellect and scholarship.

However, even assuming that professors of education are able to attract the very brightest students to work on advanced degrees, methods of teaching research methodology will have to be upgraded. One study³³ indicated that only 61 per cent of doctoral-granting institutions required work in measurement and research for a doctorate in education. One would expect an even smaller number of institutions to require work in research at the master's degree level.

Considering the need for research in education, it might be expected that professors of education would be devoting relatively large amounts of time to research. But college professors, like their counterparts in the public schools, are occupied with myriad administrative and teaching assignments, which cut rather heavily into research time. The result is that "... even in the institutions selected because of their research activities, most of the faculty members are officially assigned little or no time for research."³⁴

Adequate research requires time to think about methods of attacking important problems. In an attempt to be objective as well as to encourage research among faculty members, many universities place a heavy priority on the *number* of research papers published by faculty members rather than on the *quality* of each project. The result is that the reward system

³¹Lindley J. Stiles, "The Cooperative Research Program, Contributions and Next Steps," *Phi Delta Kappan*, XLIII, No. 6 (1962), 231-236.

³²Nicholas A. Fattu, "A Survey of Educational Research at Selected Universities," in *First Annual Phi Delta Kappa Symposium on Educational Research*, ed. Frank W. Banghart (Bloomington, Ill.: Phi Delta Kappa, Inc., 1960), pp. 1-21.

³³American Association of Colleges for Teacher Education, *The Doctorate in Education*, Conference Report, Vol. III (Washington, D. C.: The American Association of Colleges of Teacher Education, 1961), p. 43.

³⁴Fattu, "A Survey of Educational Research," p. 9.

encourages much publication which duplicates well-established research or at best makes a trivial contribution. Thus, rather than devoting valuable time to planning for an important investigation, many researchers find the need to publish so strong that the quality of their research suffers.

The lack of training in research methodology among professional educators also stems from another source. In most of the academic disciplines such as chemistry, English, or mathematics, students planning to work on advanced degrees generally have completed an undergraduate major in their subject. In addition, if these students plan on teaching in the public schools, they probably have taken at least one year's work in education primarily oriented to the acquisition of teaching certificates. Thus, if students are interested in working on an advanced degree in education, they ordinarily have undergraduate training which is stronger in their outside field of study than it is in education. For example, psychology majors usually have taken work in measurement, statistics, and research design before they enter the graduate school. Their counterparts in education who wish to specialize in educational administration, or in elementary or secondary education, begin their graduate study somewhat at a disadvantage because of their lack of undergraduate research training. This implies that the students in education usually need a rather long training program if they are to be competent researchers.

Bridging the Gap Between Theory and Practice

One factor which has retarded progress in education has been the difficulty of implementing educational research in the classroom even where theory exists. It has been estimated that the lag between the publication of research findings and the time it takes for these findings to be implemented in schools runs as high as 50 years! With this excessive lag of time, it is little wonder that educators may conclude that research can have little effect on the public schools.

Americans are pragmatic in the sense that they expect research to be judged by its products. Unfortunately, some persons have considered "products" rather narrowly and have thus been concerned only with the immediate and practical utility of research findings. Unless educational research leads directly to educational reform, they become disillusioned. One is reminded of Krutch's description of Thomas Huxley's attempts to combine the search for "truth" with utilitarian values in his support of Darwinism. Krutch stated, "He [Huxley] had received letters from well-meaning old ladies who asked what good it would do to go about assuring people that they had apes rather than angels for cousins-german—even though he himself were sure of the fact—and sometimes he felt the inconvenience of being compelled to defend his beliefs upon the double ground of their truth *and* usefulness."³⁵

³⁵Joseph Wood Krutch, *The Modern Temper* (New York: Harcourt, Brace & World, Inc., 1929), p. 39.

Educators, much like Huxley, may feel the "inconvenience" of having to increase knowledge about education as well as to provide for educational reform. These two objectives, although by no means incompatible, have tended to attract some advocates who assert that the primary function of educational research is to solve problems of a practical nature, and others who take the position that ultimately the best research comes from the contribution theory can make to the teaching-learning process.

The traditional distinction between basic and applied research was that the former was concerned with theory and the latter with practice. Thus Frank N. Freeman was simply describing the then-current status of basic and applied research when he said, "The motive in educational research is practical. . . . The motive in pure science is different. The pure scientist, as pure scientist, is not concerned with the practical consequences of his search for facts or laws."³⁶

But these stated differences between "practical" and "theoretical" may tend to obscure the interrelationships between the two activities. For example, the history of the physical sciences has indicated that what begins as a problem in theoretical research may, some years later, be of great practical importance. Thus, early workers in atomic theory did not predict the construction of the atomic bomb or the use of radiation to combat cancer. The practitioners of science—the applied scientist and the technologist—have taken science's principles and have put its products to work to satisfy man's needs and desires.

On the other hand, the "pure" or basic sciences have depended upon applied science and technology. Many industries, for example, employ large numbers of scientists to do basic research, in the hope, of course, that the knowledge obtained can be profitable. Thus, the chemical industries may employ as many scientists developing knowledge about chemistry as they do scientists whose primary responsibility is developing a salable product.

In the social sciences, the interrelationship between basic and applied research is as much of a necessity as it is in the physical sciences; educators, as social scientists, need to develop reliable knowledge about education and to test the consequences of this knowledge in schools. Considered in this manner, the study of education involves both basic and applied research, each related to the other.

Unwillingness to recognize the practicality of theory or the contributions to theory that applied research can make has led to an unfortunate split among educators. Thus one may hear "practical administrators" attacking "ivory-tower" professors for their lack of understanding of practical administrative problems. And, conversely, the educational theoretician may look with dis-

³⁶Frank N. Freeman, "Controlling Concepts in Educational Research," in *The Conceptual Structure of Educational Research*, T. R. McConnell, Douglas E. Scates, and Frank N. Freeman (Chicago: University of Chicago Press, 1942), p. 38.

dain upon the administrator for his "anti-intellectual" bias. Practical and theoretical matters should be considered as parts of a continuum with which all should be concerned, rather than as separate entities which acquire their own advocates.

An attempt to implement basic research findings in the schools by applying the principles of group dynamics to facilitate change has led to the development of what has been called *action* or *cooperative* research, which is the source of much controversy in education. At one extreme are those who claim that action research is a new methodology original to and arising from educational problems;³⁷ at the other are those who contend that action research and scientific research are incompatible.³⁸ Instead of joining the debate, it may be more instructive to consider the characteristics of action research in greater detail.

Action research in education developed in response to the need for translating educational theory into workable classroom practices. Its intent, if not always its accomplishment, was to help teachers solve some of their problems by involving them in the research process itself. It recognized that the translation of theory into practice requires the cooperation and personal involvement of teachers working closely with research specialists.

An example may help to clarify the methods, strengths, and weaknesses of action research. In one study reported by Corey,³⁹ the investigators were social studies teachers who, before the investigation was begun, tentatively accepted the hypothesis that there should be a positive correlation between the knowledge pupils have of famous characters in American history and the extent to which pupils admire these characters. The teachers stated, "This prediction is based on our belief that we should give more emphasis to the biographical method." A second hypothesis was that admiration for characters in American history should increase as a result of a one-semester course in American history. A third and final hypothesis was that there should be some relationship between the degree of admiration students feel for characters in American history and the extent to which these pupils are judged to behave in accordance with traits similar to those found in historical personages.

To obtain knowledge of the extent to which historical persons were admired, a list of 28 names was passed out to students in three classes, who were asked to check every historical character for whom some degree of admiration was felt. Students were also asked to indicate on a 5-point scale the extent of their admiration (5 points for "Admire very much" and 1 point for "Admire merely because others do"). Each student received a "Degree of

³⁷Norton L. Beach, "Research Goes into Action," *Journal of Educational Research*, XLVII, No. 5 (1954), 351-358.

³⁸Robert M. W. Travers, *An Introduction to Educational Research* (New York: The Macmillan Company, Publishers, 2nd Ed., 1964), pp. 54-56.

³⁹Stephen M. Corey, *Action Research to Improve School Practices* (New York: Bureau of Publications, Teachers College, Columbia University, 1953), pp. 61-70.

admiration score." obtained by summing the number of points given for each person admired and dividing by the number of persons admired. A sociometric measure was used to evaluate peer group relationships, and a matching test of ten items was used to measure pupils' recognition of historical personages.

For the three classes combined, the teachers found that the correlation between the amount of information about historical characters and the degree of admiration was only ± 0.05 . The mean "Degree of admiration score" at the beginning of the semester was 3.70 and at the end, 3.79; the correlation between student reputations with classmates and admiration scores was 0.07.

A few comments are in order concerning the execution of this study. In the first place, there is a great need for teachers to evaluate empirically their own beliefs and practices. As its proponents readily admit, action research cannot be used to develop general theoretical positions or to test beliefs and practices having wide applicability. However, if the results of such investigations are going to be useful to the teachers themselves, they should be aware of the extent to which currently obtained research findings are applicable to their future classes of students. In the present investigation, this factor was not evaluated.

Second, if the teachers were interested in whether or not they should give more emphasis to the biographical method, a correlation was not the appropriate statistic to use since it does not imply causality. What is needed to test the hypothesis is an experimental comparison between the effects of the biographical method and other methods which might produce important effects. Even if the correlation between amount of information about historical characters and degree of admiration were perfect ($+ 1.0$), this does not mean that "learning about these accomplishments would result in admiration."

Third, action research in general suffers from a naïve empiricism which refuses to examine what previous research has found. Even a casual search^{40, 41} through the literature would have pointed out that: 1) only about half of children's favorite characters are historical; 2) fewer than 10 per cent of children select characters of the opposite sex; and 3) children do not always identify with historical characters who have personality traits in common with their own, but may identify with those having some traits quite divergent from their own. Perhaps even more important, one investigation had already been published which showed that knowledge about historical characters and students' attitudes were largely independent of each other.⁴² Had these

⁴⁰M. Louise Stoughton and Alice M. Ray, "A Study of Children's Heroes and Ideals," *Journal of Experimental Education*, XV, No. 2 (1946), 156-160.

⁴¹David Spence Hill, "Personification of Ideals by Urban Children," *The Journal of Social Psychology*, I, No. 3 (1930), 379-392.

⁴²John R. Rackley, "The Relationship of the Study of History to Student Attitudes," *Journal of Experimental Education*, 9 (Sept. 1940), 34-36.

teachers gone to the trouble to review the literature *before* developing their hypotheses, they might not have used a list of 28 historical characters that contained the name of only one female (Harriet B. Stowe); they might have included contemporary heroes along with historical characters, and they would almost certainly have used some measure of character other than a sociometric device.

Fourth, the tests administered may have been unreliable, as the teachers themselves suggested. This is likely since the maximum "Degree of admiration score" was only 5 points. But, since the teachers did not bother to test the reliability of their instruments, there is no way of telling if the negative results were a consequence of unreliable measures, particular teaching methods employed, textbooks, teacher personality, etc. These factors were left uncontrolled.

Action research may help to reduce the lag between the discovery of knowledge and its application in the classroom. As such, it will demand more knowledge of research methods among its practitioners rather than less, and will require greater care in the dissemination and application of theoretical information by the academician.

Uncritical Acceptance of Research Findings and Methods

The uncritical acceptance of research findings is another factor that has hindered educational progress. The willingness of some persons to unquestioningly accept any conclusion that is part of a research paper has been the source of much difficulty in education. Where data have been poorly collected or where experimental designs are weak, the application of research findings in the schools must, of necessity, be misleading. This uncritical acceptance of research information has the effect of blinding the student into refusing to question existing beliefs and practices; he will not investigate what is, at least to him, a well-established principle.

Once accepted as a part of "knowledge," these beliefs are remarkably resistant to change. One author has gone so far as to call the uncritical acceptance of research findings the "standardization of error."⁴³ That is, after an idea has been circulated and generally accepted, it is almost impossible to change the attitudes of those who originally agreed with it. A celebrated case in point is the purposeful hoax perpetrated by H. L. Mencken in 1917. Writing in the New York *Evening Mail*, Mencken admonished his readers for failing to recognize "... one of the most important profane anniversaries in American history, to wit, the seventy-fifth anniversary of the introduction of the bathtub into These States." Mencken facetiously went on to describe some of the difficulties early bathtub users faced, such as attacks against the use of the bathtub by the medical profession, legislation restricting its use to certain months of the year, and the bathtub tax.

⁴³Vilhjalmur Stefansson, *Adventures in Error* (New York: Robert M. McBride and Company, 1936).

Mencken was so successful in fooling the public that in May of 1926 he felt that the hoax had gone far enough, and he admitted publicly the inconsistencies, errors, and deceptions which he had employed. But this by time the hoax had been too well accepted. It had been reported as fact in numerous professional journals, articles, and newspapers, and the error had been completely standardized. One newspaper went so far as to print Mencken's admission only to reprint his original false article three weeks later!

Researchers, whether they are in education or the physical sciences, have the obligation to evaluate research methodology and not only to question the conclusions of research, but also to be aware that conclusions are dependent upon the manner in which data have been collected and analyzed. Failure to do so may lead to the continued acceptance of incorrect conclusions.

In education, the uncritical acceptance of research findings is altogether too common. Studies and investigations are reported which may take years to discount. As an example, it is often cited as fact that a mental age of 6.5 is the minimum necessary to begin a reading program. Others have dogmatically asserted that the minimum mental age is 6.0. Neither of these conclusions can be justified. They are based upon a 1931 study by Morphett and Washburne⁴⁴ on the achievement of first-grade children in learning to read. These investigators found that almost half of the children studied who had mental ages of 6.0 or less failed to read up to the criteria established for satisfactory reading, and that the majority of children having mental ages of 6.6 or higher were able to meet these same criteria.

However, an examination of the materials required for reading in 1931 indicates that they would be considered too difficult for many of the children attending school today. In addition, the criteria for satisfactory reading were set at a higher level than most teachers today would require. The implication, of course, is that with similar materials and better teaching techniques, mental ages need not reach some magic point before reading instruction may be begun, and that what constitutes failure at one period of time is not necessarily a constant for all periods.

Reading is not the only area in which incorrect conclusions have been perpetuated. In educational measurement, for instance, numerous examples can be found of erroneous beliefs existing long after they were shown to be incorrect. Thus, many teachers still continue to compare achievement scores with IQ scores for the purpose of determining "under- and overachievers." This practice began as early as 1920, when Franzen popularized the use of the so-called "accomplishment quotient," a ratio of achievement test scores to intelligence test scores. A ratio of 100 meant that the child was doing as well as could be expected, a ratio of less than 100 meant he was underachieving and a ratio above 100 meant overachievement. It has been shown repeatedly

⁴⁴Mabel V. Morphett and C. Washburne, "When Should Children Begin to Read?" *The Elementary School Journal*, XXXI, No. 7 (1931), 496-503.

that the use of such a ratio violates basic and necessary assumptions about the meaning of educational and psychological measurements. Indeed, Franzen himself later repudiated its use. Nevertheless, teachers continue making the comparison even though there are sound arguments against doing so. To make things worse, many modern textbooks in education still continue the practice of making direct comparisons between mental age scores and such measures as reading or arithmetic age scores. What first appeared to have much educational merit later turned out to have enjoyed a use far beyond its values.

SUMMARY

1. Educational research has been characterized as showing great promise but few accomplishments. In part, the failure of educational research to live up to the high expectations of educators has resulted from the contradictory goals proposed for education and from the inability of educators to implement research findings directly into the classroom. Whether or not research findings will be implemented depends upon their relationship to social, economic, and political factors operating in the community.

2. A distinction was made between *basic* and *applied* research. The purpose of basic research is to develop theory; applied research resolves immediate and practical problems. The point was made that theory can suggest what variables are important in undertaking an applied research problem. Empirical and theoretical studies are needed if applied research is not to become overly repetitious and trivial.

3. Various types of research activities and their relationship to educational innovation were described:

a. *Development projects* are used to increase efficiency and to lower operating costs so that proposed innovations can be made practicable. Development depends upon knowledge from both theoretical and empirical studies to help justify the time and effort expended on a proposed project.

b. The *demonstration* is used to persuade teachers and administrators that a proposed innovation is useful and desirable. It depends upon development projects to make proposals workable, but in turn it may also create new research problems as operating deficiencies and weaknesses are field tested.

c. Three categories of research were described which contribute to both basic and applied problems:

(1) The purpose of *analytic research* is to derive relationships deductively. It helps to point out implications and possible consequences of proposed innovations and helps to clarify assumptions behind proposed studies. It uses the language of history, philosophy, mathematics, and linguistics. *Data retrieval* was described as a form of analytic research designed to gather information from both primary (original) and secondary (non-original) sources.

(2) *Descriptive research* involves the use of correlations, surveys, case studies, direct observational techniques, cross-cultural methods, growth studies, etc., which are designed to describe existing conditions without their being influenced by the investigator. During the initial stages of an investigation, descriptive research methods help point out the extent and current status of a problem.

(3) *Experimentation* is used to test causal relationships by holding extraneous conditions constant to determine the effect one variable has upon another. An experimental design is especially applicable in comparing the effects of two or more innovations. Experimentation often follows analytic and descriptive investigations.

4. Historically, educational research developed from the need to modify the curriculum and to evaluate various educational innovations. The motive for research was practical and utilitarian, with theory being given little consideration. Much of educational research today still suffers from a lack of theory and from an excessive number of fragmentary research studies which are not organized into theoretical positions.

5. Few empirical studies in education were conducted prior to the twentieth century. Educational research awaited the development of psychological techniques and methods (especially in testing) before educational problems were studied scientifically. Herbart (1776-1841) is generally given credit for suggesting that it was possible to study education scientifically, if not experimentally. It was Wundt's laboratory (established in 1879) that gave impetus to experimental methods in the social sciences. William James's laboratory at Harvard (1875) and G. Stanley Hall's at Johns Hopkins (1883) furthered experimentation in the behavioral and social sciences.

Table 2 summarizes a number of other important nineteenth-century contributions to educational research.

6. The rise of the psychological "schools" did much to encourage research in education. By emphasizing the values of introspection in studying the content of the mind, *structuralists* such as Wundt and Titchener encouraged the study of consciousness and awareness which modern techniques of interviewing and factor analysis have continued. The *functionalists* rose in opposition to the structuralists. As a functionalist, Dewey was more concerned with the purpose of the mind than with the analysis of its contents. As the most deliberate spokesman for progressive education, Dewey did much to modify the sterile and impractical education demanded by the memoriter system and the theory of faculty psychology.

Behaviorism rose to challenge both structuralism and functionalism. John Watson, who developed and popularized the movement, was concerned with the study of behavior, not mental functions or the content of the mind. Watson denied the concept of innate intelligence and saw the conditioned reflex as the means by which all organisms learn. Thorndike's studies of human learning did much to further experimentation into human behavior, although he never formally joined the ranks of the behaviorists.

The *Gestalt* school was founded by Köhler, Koffka, and Wertheimer in opposition to behaviorism. The Gestaltists emphasized molar analyses of

behavior and experience in contrast to the behaviorists, who insisted on a more molecular analysis of behavior. The arguments which raged between behaviorists and Gestaltists did much to develop new theories and to stimulate research.

TABLE 2

A SUMMARY OF SELECTED NINETEENTH-CENTURY CONTRIBUTIONS TO EDUCATIONAL RESEARCH

<i>Event</i>	<i>Year</i>	<i>Responsible Person</i>	<i>Importance</i>
<i>Origin of Species</i> published	1859	Charles Darwin	Showed that human beings could be studied and that they obey the same laws of nature as do other animals
<i>Expression of the Emotions in Man and Animals</i> published	1872	Charles Darwin	Showed that emotion could be studied in humans
First testing center developed	1882	Sir Francis Galton	Helped popularize testing and the measurement of individual differences
Development of the questionnaire	1883	Sir Francis Galton	Introduced possibility of studying large numbers of persons at a given time
First experimental analysis of learning	1885	Hermann Ebbinghaus	Developed the nonsense syllable and various techniques for measuring learning and retention
<i>Pedagogical Seminary</i> published	1891	G. Stanley Hall	Helped disseminate information on child growth and development
<i>Animal Intelligence</i> published	1898	E. L. Thorndike	Gave impetus to animal experimentation
<i>Talks to Teachers</i> published	1899	William James	Popularized educational psychology

The *psychoanalytic* movement was begun by Freud in an effort to explain personality, especially abnormal personality. By emphasizing the study of individuals, Freud gave impetus to case study methods.

7. The psychological schools were unable to continue unchanged for very long. In part, this was due to 1) rigid demands required by the early founders of the schools, which tended to alienate and create splinter organizations; and 2) the amalgamation of various positions as new techniques and methods of studying human behavior were discovered.

8. Research in educational administration developed out of social needs to reform the public schools, in the same way that the public demanded reform in industry and government. The goal of nineteenth-century and early twentieth-century school administrators was greater efficiency of operations.

The purpose of research was to help administrators reduce costs and increase school services.

A number of professional organizations in educational administration have done much to develop a science of administration. With the help of the Kellogg Foundation, the American Association of School Administrators, and various organizations of professors of administration, research centers have been established to study administrative problems and to develop and test theories of administrative behavior.

9. The impetus for research on the curriculum developed out of practical needs. By the twentieth century, educators were faced with large numbers of students who found the prevailing curriculum rigid, difficult, unmotivating, and unrelated to "real-life" conditions. The criterion of *social usefulness* was used to determine curricular content. This led to surveys of activities engaged in by members of various occupations and professional groups.

The need for curricular reform encouraged the use and evaluation of innovations in the public schools. Unfortunately, these innovations have not been systematically or thoroughly tested before they have been recommended for general use. Nonetheless, there have been a number of theories of curriculum and instruction proposed within the past few years, including those of Gagné, Piaget, Guilford, Bruner, Carroll, Suppes, and Suchman, which are being empirically studied and evaluated.

10. Although the use of research techniques has helped improve public education, a number of factors have hindered the development of educational research.

a. *Expenditure of funds.* It was stated that expenditures for educational research are inadequate to do the job which educators have demanded of research.

b. *Lack of training.* Another factor which has hindered the development of educational research is the lack of training given to most professional educators in research methods. It was indicated that even where adequate training in research has been provided, time factors and the "publication race" often interfere with the better judgment of researchers.

c. *Inability to relate theory and practice.* One factor retarding the development of educational research has been the inability of educators to implement research findings. The traditional distinction between basic and applied research was modified so that the interrelationships between them could be emphasized. This led to a description and analysis of *action research* as an attempt to relate theory and practice.

d. *Uncritical acceptance of research finds.* Because most teachers have been inadequately prepared to understand research findings and methods, they are likely to be uncritical consumers of educational research. Many of these uncritically accepted ideas are highly resistant to change.

PRACTICE EXERCISES

1. Suppose that you are the superintendent of a small city school district and that you have a number of research findings and recommendations on your

desk. Indicate which of these recommendations you would be willing to implement in your schools, assuming, of course, that the findings are reliable. Indicate what factors led to your decision either to put these recommendations into operation or to refuse to implement them.

Findings

- a. Girls are more mature than boys at comparable chronological ages.
- b. High school students who take courses in driver education have fewer accidents than those who are not given such training.
- c. Students who retake courses after failing them do less well when they retake the course than do their counterparts who were passed.
- d. The correlation between teaching "success" and the amount of knowledge obtained in college courses yields low but positive correlations.
- e. In communities where playgrounds are made available, there is less juvenile delinquency than in communities which do not provide playgrounds.

Recommendations

- a. Girls should enter school a year earlier than boys.
- b. Schools should provide courses in driver education.
- c. Students should not be failed.
- d. Administrators should select teachers more on their personality and character than on their achievement.
- e. Administrators should allow students to use school playgrounds after school.

2. Classify each of the following examples into the type of research or activity it calls for (basic, analytic, descriptive, data retrieval, development, or demonstration):

- a. Two sets of data are correlated.
- b. A teacher looks up the research on the teaching of reading.
- c. A teacher tries to show parents the advantages of a new textbook in mathematics.
- d. An old course of study is being redone to make it simpler and more palatable to students.
- e. A student is interested in knowing the number of children who attended summer school during the 1965 summer session.
- f. A teacher believes that stuttering is caused by neurological problems.
- g. An experimenter reviews the literature to find evidence for some position.
- h. A business teacher installs a time clock in his classroom to see if this improves the attitudes of students toward being efficient.

3. Examine in detail any of the following new programs. What assumptions does the method make about human learning? What evidence is there that the assumptions are correct? What other methods could be used that might bring about the same objectives?

- a. Team teaching.
- b. Ungraded classrooms.

- c. Individualized reading programs.
- d. Chemical Bond Approach.
- e. UICSM math program.

SUPPLEMENTARY READINGS

1. Abelson, Philip H., "Congress and Research," *Science*, 139, No. 3552 (Jan. 25, 1963), 305. This article indicates the need for greater amounts of money for educational research. The author points out some of the dangers involved in the acceptance of research grants.
2. Barr, A. S., "Research Methods," in *Encyclopedia of Educational Research*, ed. Chester W. Harris (3rd ed.) (New York: The Macmillan Company, Publishers, 1960), pp. 1162-1166. Part of this chapter contains a brief but well-documented history of educational research organized into three parts: Educational Research Before 1900; Educational Research, 1900 to 1925; Educational Research After 1925.
3. Boring, Edwin G., *A History of Experimental Psychology*. New York: Appleton-Century-Crofts, 1950, 777 pages. Although not specifically related to the historical development of educational research, this text does describe in detail the rise of science and the development of psychology as a scientific enterprise. The origin of modern psychology within philosophy is treated in Chapters 9-13, the contributions of Wundt are described in Chapter 16, and Chapters 21 and 22 are concerned with the establishment of psychology in America.
4. Lazarsfeld, Paul L., and Sam Sieber, *Organizing Educational Research*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1964, 113 pages. A short but excellent account of the relationships among the various organizational structures within educational research bureaus. Of special relevance to this chapter is Chapter 3, "Problems of Educational Research and Research Institutes."
5. Miles, Matthew B., *Innovations in Education*. New York: Bureau of Publications, Teachers College, Columbia University, 1964, 689 pages. An excellent analysis of the factors which retard and advance educational progress.
6. Morse, Arthur D., *Schools of Tomorrow—Today*. New York: Doubleday & Company, Inc., 1960, 191 pages. This is a progress report on various educational experiments such as the use of school aides, team teaching, and television. The selections are nontechnical and are designed to provide the reader with an overview of new approaches to persisting educational problems.
7. Postman, Leo, ed., *Psychology in the Making: Histories of Selected Research Problems*. New York: Alfred A. Knopf, Inc., 1962, 785 pages. Students will find the introduction to these readings of interest. Although specifically titled "Some Guides to the Understanding of the History of Psychology," this chapter is an attempt to put research into a cultural

context and, as such, it provides an excellent framework for research in disciplines other than psychology.

8. Stanley, Julian C., "Doctoral Students and Research in Educational Psychology," *School and Society*, (Nov, 3, 1962), pp. 375-377. The author outlines a training program for educational psychologists which emphasizes high standards of selection and training. His plan, if followed, should do much to alleviate the lack of training in research methods. The article, while restricted to the training of educational psychologists, could very well be applied to the training of specialists in other areas of education.

CHAPTER THREE

Selecting a Research Problem

The student who is attempting to formulate his first research project often finds himself in a bewildering situation. He may have repeatedly come across statements from his instructors and from textbooks recommending additional research to help clarify the many pressing problems in education. Thus, often armed with the loftiest ideals, he begins his search for a problem worthy of investigation. Almost immediately, however, his project may be criticized as being unimportant, unnecessary, too impractical, or already "worked to death" by other investigators.

One reason students find it so difficult to select a research problem is that they often confuse *problems* with *methods*. A problem is "a perplexing situation after it has been translated into a question or series of questions that help to determine the direction of subsequent inquiry."¹ In contrast, a method is a procedure used in resolving problems. To begin by considering method without having a clear formulation of the problem is to invite trouble. The student who proposes to investigate the correlation between intelligence and achievement test scores, for example, is concerning himself with a method. What he has failed to indicate is the *need* for an investigation on this topic.

Research problems consist of questions that can be justified on grounds to be discussed later in this chapter. The process of justifying questions is extremely important, because it implies that not all problems are of equal value or importance. It is the responsibility of the researcher to demonstrate the need for a particular type of investigation to justify spending time, effort, and funds on one study rather than another.

The necessity of justifying research problems stems from a number of sources. In the first place, scientific endeavors are public matters; as such they command the interest and efforts of the scientific community at large.

¹Carter V. Good, ed., *Dictionary of Education* (2nd ed.) (New York: McGraw-Hill Book Company, 1959), p. 414.

The investigation of a strictly personal problem having applicability to oneself alone is not likely to contribute to a science of education.

Second, it is not possible for researchers to investigate all of the many problems educators face. Thus, it is necessary to be selective about those problems which are to receive priority. Decisions concerning these priority judgments require criteria for their evaluation. A number of criteria have been proposed for justifying the selection of a research problem.

The Criterion of Interest

The criterion of *interest* is a necessary but not sufficient reason for selecting a particular research problem to investigate. Without interest in a topic, the effort and care characteristic of well-designed research studies are likely to be lacking. Interest is not, however, a *sufficient* criterion, since it can lead researchers to investigate trivial problems as well as important ones. As we have already pointed out, the nature of scientific research demands that findings be communicated, and little is gained by communicating matters that do not concern other scientists. This by no means requires the researcher to select a problem which is of concern to everyone. It is the *probability* of a study making a contribution that determines its value to other scientists, but the researcher is not obligated to select a problem that is first on the priority list. A study designed to count the rocks on a playground may provide us with information and may even be of interest to the person doing the counting, but the probability of its being of interest to anyone else is low.

Interest does not necessarily imply agreement with a given state of affairs. That is, one does not have to be in agreement with a body of knowledge in order to investigate some aspect of it. Many of the most important investigations have been performed as reactions against currently accepted beliefs. This type of research is important because it points out weaknesses in previously conducted studies and stimulates research on new topics. Accepted positions can be differentiated, investigated, and reformulated in the light of new research findings that stir up controversy.

The Economic Criterion

Another criterion for selecting one research project over another is *economic*. The realities of life may force the researcher to reject an investigation if his estimates of cost are larger than his available funds; costs also may force him to limit the breadth or complexity of a proposed study. While hardly a justification for undertaking a research project, costs related to the hiring of personnel, the use of computing facilities, the construction or purchase of equipment and materials, and overhead must be considered, especially in "applied" research, where some ratio of cost per yield is often required. When the researcher expects some sort of economic return for his expenditures (as, for example, in industry), then costs can be computed. In "pure" or basic research, where the contribution is primarily theoretical, the yield may be impossible to estimate.

There is, of course, no justification whatsoever for undertaking a project simply because it is inexpensive or easy to execute. Like the criterion of interest, costs must be considered in any investigation, but do not in and of themselves provide sufficient justification for selecting a research problem.

The Investigator's Ability and Training

Another factor to be considered in the selection of a research problem is the *investigator's ability and training*. Every person embarking upon a research career has limitations in experience and capacity which he should recognize if he is to avoid unnecessary frustration. It is at this point that intellectual honesty with oneself is essential. Not everyone is capable of undertaking complex research problems; it is a sign of maturity to be able to recognize one's own limitations.

Even if an investigator is *potentially* capable of undertaking an elaborate research study, there is still the matter of his training and the amount of time he may be willing or able to devote to the project. The design of some studies is almost certain to require the investigator to have competence in areas where he has not had formal training. In some instances, this may be remedied rather quickly by reading a number of articles or texts or by asking someone who is competent to help on the project in an advisory capacity. On the other hand, taking a course in projective techniques does not by itself lead to competency in projective testing. It may take a great deal of practice under supervision to interpret some tests. The student should be able to evaluate his strengths and weaknesses before he undertakes a research project. To wait until the project has begun is to invite needless delay at best and an incomplete or meaningless project at worst.

The Criterion of Uniqueness: the Problem of Replication and Generalization

The criterion of *uniqueness* assumes that research projects have some originality in either purpose or method. It is most unlikely, of course, that the novice researcher will come across a problem never before considered; we will concern ourselves here with the question of the advisability of duplicating research that has already been completed.

Suppose, for example, that a fifth-grade teacher decides to evaluate the effectiveness of a new method of teaching fractions by experimenting with the students in her class. Half of her students are given the new method and the other half are taught by her usual procedure. Now if the group taught by the new method actually did learn fractions "better" than the group taught by conventional methods, what could this teacher legitimately conclude? Certainly she could not generalize from her findings and recommend the new procedure unquestioningly for all fifth-grade teachers or for all fifth-grade children. Because other fifth-grade teachers and classes were not used, her findings have limited application and would therefore be of less generality than a study which involved numerous classes and numerous teachers. A

more comprehensive study is required. It would be perfectly reasonable for a researcher to take a study that was originally limited in scope to try to make it more comprehensive and, ultimately, of wider applicability.

On the other hand, not every study which involves some event needs to be *replicated*, or repeated, using various groups, if the conditions which underlie the event are made explicit. Thus, fairly unique events may be investigated if the underlying theoretical conditions are specified. For example, on October 30, 1938, the originally scheduled radio program on the Mercury Theatre was "interrupted" by the announcement that Martians had invaded the United States. Almost immediately, panic broke out among the program listeners, many of whom believed the story of the invasion to be true. As one investigator reported:

Long before the broadcast had ended, people all over the United States were praying, crying, fleeing frantically to escape death from the Martians. . . . Such rare occurrences provide opportunities for the social scientist to study mass behavior. . . . The situation created by the broadcast was one which shows us how the common man reacts in a time of stress and strain. . . . The panic situation we have investigated had all the flavor of everyday life and, at the same time, provided a semi-experimental condition for research. In spite of the unique conditions giving rise to this particular panic, *the writer has attempted to indicate throughout the study the pattern of the circumstances which, from a psychological point of view, might make this the prototype of any panic.*²

In the study of education, occasions arise when nonreplicable, unique events may be investigated. The process of school integration, for example, would be difficult, if not impossible, to study in other contexts or under other conditions. Thus, a unique event may be studied as long as the conditions are stated which make it a model for other related and similar events.

In another sense, uniqueness as a criterion for the selection of a research problem does not require that every event be studied in every context. This relates to the problem of just how far research findings can be generalized. Usually, the most useful investigations are those which allow the researcher to make the broadest and most inclusive generalizations concerning any given topic; conversely, investigations which are of limited applicability, either theoretically or practically, are of less value. Thus, a hierarchy of problems can be formed, with the most inclusive one having the highest priority. In practice this would mean that, given a choice of possible problems to be studied, the student would be making the greatest contribution by selecting the problem that has the greatest probability of yielding results which lead to general theories and laws.

²Hadley Cantril, "The Invasion from Mars," in *Readings in Social Psychology*, (3rd ed.), ed. Eleanor E. Maccoby, Theodore M. Newcomb, and Eugene L. Hartley (New York: Holt, Rinehart & Winston, Inc., 1958), pp. 291-300. Italics added. Reprinted by permission of Princeton University Press.

RESEARCH TO CLARIFY AND VALIDATE BASIC THEORETICAL PROBLEMS

In Chapter One we defined a theory as a unified system of definitions, postulates, and observations organized in such a way as to most simply explain the interrelationships between variables. We also pointed out that while theories themselves cannot be investigated directly, there are empirical consequences predictable from theory which can be validated. If these consequences are found to be in the predicted direction, they lend credence to the theory; if not, the theory will have to be modified to account for exceptions.

There are two ways in which theory can aid in the selection of a research problem. First, research projects can be undertaken to help clarify, limit, and define theoretical formulations. Take, for example, the interference theory of forgetting described in Chapter One. Somewhat oversimplified, it states that forgetting occurs because newly acquired information is interfered with by both previous and subsequent learning activities. This brief statement of the theory, however, leaves many questions unanswered. For example, what is the relationship between number of previous (or subsequent) learning activities and amount of forgetting? Do previously learned activities interfere more than do subsequently learned activities? Do all activities interfere with newly acquired information or can some activities facilitate retention? Each of these questions forms the basis of a research problem which can be justified on the grounds that the research will help to clarify the meaning of the theory.

Second, the consequences of theory can be used to test indirectly the validity of the theory itself. The interference theory, for example, *implies* that: 1) there should be a positive correlation between amount learned and amount of forgetting; 2) material located at the ends of lists to be learned should be retained better than material located at the center of the list; and 3) not all activities interfere with retention since, for example, previous knowledge of how to play the violin may facilitate learning the viola. Each of these implications provides a basis for a research problem.

The relationship between empirical investigations and theory can be seen by examining Figure 1, where the *EI's* represent *empirical investigations* designed to help develop a body of theory. Each empirical investigation adds further knowledge to the development of a theory. In turn, the theory implies certain consequences (*C*), whose validation lends credence to the theory. In the beginning stages of theory development, the empirical investigations provide information about the theory; once the theory is better understood, the consequences of the theory are tested to determine if they are in accord with what the theory predicts. In the selection of a research problem, investigations may be chosen at either the *EI* level or the *C* level. In either case, the student should be familiar with the major empirical investigations

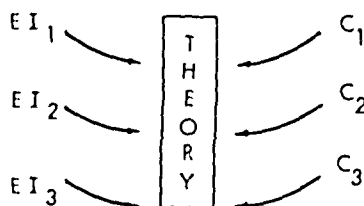


Fig. 1

The relationship between empirical investigations and theory development.

leading to the formulation of the theory and to studies which have tested the consequences of the theory.

RESEARCH TO CORRECT METHODOLOGICAL ERRORS

Research to Clarify Contradictory Findings

The principle of determinism requires that investigators examine research *methods* for possible discrepancies if there are contradictory conclusions from research investigations. Where experimental designs have been weak, conclusions based upon these designs are open to serious question.

An example will help demonstrate the point. In discussing *underachievement*, Peterson has indicated a number of research findings which seem to contradict each other:

Consider these conclusions, for example. It has been observed that the father of the low achiever has less education and ranks lower in occupational status than the father of the high achiever. Yet, another study states the opposite—*more* education and *higher* occupational levels among the fathers of low achievers. And a third study finds no difference in educational or occupational level between the two groups.

Helpful? Let's look further: The home of the underachieving student is more likely to be broken by death or divorce, or, the loss of one or both parents has no effect on achievement. Take your choice.

Is birth order important? One author states that the achievement of the only child or the oldest sibling is poorest. Another reports no significant differences between older and younger siblings in intelligence or achievement.³

In the example just cited, an apparent contradiction in findings could lead a researcher to design studies which might provide more definitive answers concerning the characteristics of the underachiever. In the first place, the definitions of "underachievement" were different in each of these studies. Peterson himself has indicated that the use of four different criteria

³John Peterson, "The Researcher and the Underachiever: Never the Twain Shall Meet," *Phi Delta Kappan*, XLIV, No. 8 (1963), 379.

for determining what shall be considered underachieving behavior probably contributed to the wide discrepancy in reported findings. In some studies, underachievers may be from widely different types of schools or from different communities, and may vary considerably in age and grade. In addition, the estimation that the child is able to do better work than he is actually doing may come from different sources. In some instances the source may be the teacher's judgment; in other cases, it may be standardized test results. Because the characteristics of underachievers are determined by such diverse sources, each having different criteria as to what constitutes underachievement, studies are likely to show contradictory findings.

Just how could one select a problem which attempts to clarify contradictory findings? Using studies on underachievement as an example, the researcher could repeat earlier contradictory investigations, selecting a number of different types of schools in different communities, using elementary and high school children of both sexes, and specifying his criteria for determining underachievement. To be sure, this would lead to a rather complex study, but it would help to clarify findings which now appear to be inconsistent.

Research to Correct Faulty Methodology

Another source for selecting a research topic may be found in generally accepted conclusions which are based on faulty research designs. In educational and psychological measurement, for example, it is generally well accepted that scores which are corrected for guessing tend to correlate very highly with uncorrected test scores. This conclusion is based upon the following type of research design:

1. One group of students takes a test.
2. The papers are scored twice, once by correcting for guessing and once without the correction.
3. The correlation between corrected and uncorrected scores is computed.

The results invariably indicate a very high correlation, and this evidence is used to argue against correcting examination papers for guessing. The reason given is that since the scores are highly correlated, they must be measuring the same factors. However, Davis has indicated that the research design is faulty:

The writer discounts statements that the correlations between sets of test scores obtained with and without correction for chance are exceedingly high. In the first place, the correlations cited are usually spuriously high because they are obtained by scoring the same set of papers in two ways. In this situation, the directions for administering the tests are the same and thus can be appropriate only to *one* of the two scoring procedures. It would be desirable to obtain the correlation between two comparable forms of the same test administered successively to the same sample with two sets of directions—one appropriate to scoring without correction for

chance success and one to scoring with it. This correlation should be compared with a parallel-forms reliability coefficient [see Chapter Seven] for the same test based on the same sample. An exact test of the significance of the difference between the two coefficients would permit inferences to be drawn regarding the point at issue.⁴

Numerous other examples can be found in the literature where faulty research procedures have stimulated investigators to correct and improve upon designs. For example, in the original construction of the *F* scale (a measure of authoritarianism), all items were so worded that agreement signified hostility, ethnocentrism, and intolerance. However, Messick and Jackson⁵ pointed out that persons taking the test who had a general tendency to agree to any set of statements regardless of content would be considered to be undemocratic although they could simply be displaying their tendency to acquiesce or agree. The authors indicated that a number of studies have been made to investigate and to reduce this type of tendency, or *response set* on the *F* scale.

In general, many research topics suggest themselves after a careful analysis and critique is made of previously conducted studies. The attitude to guard against is being overly critical on minor or unimportant details. To criticize and replicate a study because of some unimportant detail adds little to knowledge.

Faulty Use of Statistics

Another type of faulty methodology involves the use of inappropriate statistics for analyzing data. Often an important study yields questionable results because the statistics used are not applicable. These studies may be replicated using more appropriate statistical methods.

An example or two may be helpful at this point. In early investigations involving the relationship between cigarette smoking and the prevalence of lung cancer, researchers computed correlations between the number of cigarettes smoked per day and the prevalence of cancer. What the researchers really wanted was evidence as to whether or not smoking *caused* lung cancer; what they actually obtained was the correlation, or degree of relationship, between smoking and cancer. This information, while useful, could not be used as decisive evidence showing a causal relationship. The researchers failed to recognize that correlations do not necessarily imply causality.

On a much broader level, the million-dollar PROJECT TALENT research sponsored by the Cooperative Research Branch of the U. S. Office of Education has been criticized on essentially the same grounds. As Stanley has stated:

⁴Frederick B. Davis, "Item Analysis Techniques," in *Educational Measurement*, ed. E. F. Lundquist (Washington, D. C.: The American Council on Education, 1951), p. 277.

⁵Samuel Messick and Douglas N. Jackson, "The Measurement of Authoritarian Attitudes," *Educational and Psychological Measurement*, XVIII, No. 2 (1958), 241-253.

I admire the testing operation itself. The American Institute for Research and the University of Pittsburgh have good cause to be proud of it, despite a few inevitable flaws in design and execution. It is the promised attempt to tease out "causation" from essentially correlational data that gives me pause. . . .

Viewed narrowly, all ten questions [asked by the PROJECT TALENT staff] deal only with association, not causation, but straightway they have causal implications that might be misleading. Take No. 3, for example: "Are small classes more effective than large classes?" One can compute quite readily the correlation of class size as it occurs "naturally" in PROJECT TALENT schools with various test scores, but it is a far more difficult task to prove that class size helps "cause" the differences among means of classes and that therefore classes "should" be of a certain size or sizes in order to maximize desired outcomes of instruction.⁶

The use of an inappropriate statistic is not an uncommon occurrence even in published research reports. Thus Lewis and Burke⁷ have stated:

To confirm a general impression that the number of misuses of χ^2 [chi-square] has become surprisingly large, a careful survey was made of all papers published in the *Journal of Experimental Psychology* during the three years 1944, 1945, and 1946. Fourteen papers were found which contained one or more applications of the chi-square test. The applications in only three of these papers . . . were judged to be acceptable. On one other paper . . . the several applications could be called "correct in principle" but they involved extremely small theoretical frequencies. In nine of the fourteen papers . . . the applications were clearly unwarranted. In the remaining case, it was not possible to determine what had been done; and the author, when questioned twice by letter, did not choose to reply.^{8,9}

If the research or article is judged to have some theoretical or practical importance, then it could very well be replicated using a more appropriate statistical technique.

RESEARCH TO RESOLVE PRACTICAL FIELD PROBLEMS

Many important research projects develop out of attempts to solve practical problems faced by classroom teachers and administrators. These problems are called practical, not because they lack theoretical bases, but because the intent of the investigator is to study a pressing and immediate problem. Thus, for example, a classroom teacher might be concerned about the number

⁶Julian C. Stanley (Book Review), *Teachers College Record*, LXIV, No. 5 (1963), 433-434.

⁷Don Lewis and C. J. Burke, "The Use and Misuse of the Chi-Square Test," *Psychological Bulletin*, XLVI, No. 6 (1949), 433-489.

⁸Lewis and Burke, "The Use and Misuse of the Chi-Square Test," p. 433.

⁹For a rejoinder to the Lewis and Burke article and their defense see *Psychological Bulletin*, XLVII, No. 4 (1950), 331-355.

of spelling words or arithmetic problems her students can effectively learn during an hour period of time. Her interest is only in her students and in solving a problem which faces her and which demands immediate attention. Similarly, an administrator may want to utilize his staff most effectively to direct extracurricular activities in his school. He may very well recognize that his findings might apply to other schools and other administrators, but his primary intent is to solve *his* problems rather than similar problems faced by other administrators.

In a sense, the problems faced by the administrator and the classroom teacher have implications that extend beyond the immediate ones. The teacher will no doubt apply current findings on spelling and arithmetic to her next year's classes. The administrator who finds that his staff was most effective when allowed to choose its own extracurricular activities to supervise will probably encourage new staff members to select their own supervisory functions in future years. The assumption which they both make is that future classes and future members of the staff will act in the same way as the subjects did when the study was undertaken. However, this conclusion is valid only if the characteristics of the students or teachers remain constant over time. The teacher and administrator will have to repeat their investigations every year to be sure that the characteristics of their subjects are not changing.

Thus, even where the intent of research is to help resolve practical difficulties, the researcher still has the responsibility for determining the limits to which the study applies and for designing a well-controlled study. A teacher who finds that ten minutes per day devoted to spelling yields optimum benefits will have to remember that these findings hold only for the kinds of students who are in her classes the year the study is made. Next year's students may be either brighter or duller than her present group.

Would it not be more practical, then, for the teacher to know the minimum periods of time required for spelling at various levels of student abilities and aptitudes than it would be to know that this year's students require only ten minutes of spelling instruction per day? And once having this information would it not be of use to other teachers? The point is that research which has limited application is limited not only for other teachers but for the teacher-researcher himself. By making the research more comprehensive, the teacher makes it more practical for himself; at the same time he can contribute materially to the benefit of his colleagues. Research which is highly restricted to one given condition (i.e., spelling) in one given classroom is not only unlikely to arouse wide interest, but may not actually be of great value to the researcher.

Somehow many persons have the attitude that if the research is to solve a practical problem, a type of research which is well controlled and which involves a statistical analysis of data need not be applied. Nothing could be farther from the truth. As Massialas and Smith point out:

At the fingertips of the classroom teacher lie innumerable opportunities for research which may be the source of additional insight into many of the problems of understanding the teaching-learning process A systematic, thoughtful, and vigorous approach to the investigation is necessary if we are to expect a clear design and valid strategy of attack. Only when tight and organized frames of reference are consciously employed will the field of education move toward quality research.¹⁰

PUBLISHED SUGGESTIONS FOR SELECTING RESEARCH TOPICS

From time to time, suggestions for selecting research topics have been published; they may be of some help to graduate students as they begin their search for adequate topics. Many of these suggestions, while worthwhile in general, may not be suitable for any given student. It should be kept in mind that the interests and capabilities of the researcher are important criteria for the selection of a research topic. In addition, a number of local considerations such as equipment, subjects, and the requirements of various graduate departments may not allow the student to pursue a given research topic. Suggested topics, may, however, provide the student with some idea as to what types of studies are needed. As suggestions, they are by no means prescriptive.

Suggestions for Research in Measurement

One excellent set of suggestions about possible research topics for graduate students has been published by the Educational Testing Service.¹¹ Although the list does not contain specific methods which could be used in attacking these topics, the student may find some useful ideas in the twelve suggestions which follow:

1. Structure in the essay question as it affects the scoring reliability and the test reliability.
2. Determining characteristics of test items at the time of writing.
3. Comparison of group tests vs. individual tests of intelligence or academic ability.
4. The effects of reducing the administration time of tests.
5. The relative values of r biserial and D as guides to item selection.
6. The relative merits of entrance test scores and first year grade point averages as bases for predicting college success.
7. The relative merits of status measures as predictors of success in high school.
8. Effects of different teaching methods used in tests and measurement courses for students in education.

¹⁰Byron G. Massialas and Frederick R. Smith, "Quality Research—A Goal for Every Teacher," *Phi Delta Kappan*, XLIII, No. 6 (1962), 256.

¹¹Evaluation and Advisory Service, *List of Suggested Studies for Graduate Students in Tests and Measurement* (Princeton, N. J.: Educational Testing Service, 1960).

9. The relationship between creativity and scores on general ability tests.
10. The effects of various mechanical means used by the student in recording test responses.
11. Prediction of properties of test scores for multiple choice items.
12. The possibility of finding groups of test items which are "reproducible."

Suggestions for Research on Gifted Children

The Division of School Psychologists of the American Psychological Association has published a list of research topics concerning the education and psychology of the gifted child.¹² In general, these suggestions fall into 11 categories: administrative procedures related to grouping, enrichment, etc.; organization and amount of schoolwork and its effects on personality and intelligence; the relationship between ability and performance; the teacher's role in educating gifted children; the vocational aspects of educating gifted children; personal relationships of gifted children to others of lesser and equal abilities; the special frustrations and problems faced by gifted youngsters; the special satisfactions of being gifted; the determination of factors responsible for undesirable personality traits in gifted children; the status of gifted children as a possible minority group; and the needs of gifted children with regard to the development of desirable citizenship traits. Many of these suggestions could also be made applicable to retarded students or to students having emotional or physical problems.

Suggestions for Research on the Curriculum

A number of persons have suggested needed research concerning the curriculum. Passow, for example, has stated:

Research is needed for developing and testing conceptual frameworks and theories that underlie curriculum building. . . . Needed are more comprehensive curriculum-research projects—an investigation for example, of the total educational program as it affects the learning patterns of youngsters. Among a few researchers there has been the tendency to analyze a thin slice of some overriding problem. Comprehensive research plans dealing with basic educational problems have not been undertaken in sufficient numbers Also needed are more longitudinal studies.¹³

Clymer has also suggested some needed research :

1. The need to evaluate present status of research findings. What has research really uncovered thus far?
2. The need to develop a theoretical framework into which all present knowledge will fit. What are the theoretical principles that serve to unify this knowledge?
3. The need to improve certain characteristics of educational research. What statistical and design factors of research need improvement?¹⁴

¹²Gertrude Hildreth, George Myer, Lee Meyerson, Paul Witty, and Harriet E. O'Shea, "Needed Research on Gifted Children," *American Psychologist*, IX, No. 2 (1954), 77-78.

¹³Harry Passow, "Curriculum Research: Status, Needs, and Prospects," *Educational Research Bulletin*, XXXIX, No. 8 (1960), 197-205, 224.

¹⁴Theodore Clymer, "Some Current Needs in Educational Research," *Phi Delta Kappan*, XL, No. 6 (1959), 253-257.

Suggestions for research have also been published in specific areas of the curriculum. Gray,¹⁵ for example, has suggested a number of possible studies concerned with the improvement of reading. These include analysis of those aspects of intelligence most responsible for learning to read, determination of limits in both reading speed and comprehension for students of various levels of ability, determination of the optimum methods for learning to read, differences in reading approaches for boys and girls, investigation of those emotional factors that retard reading instruction, and various methods of analyzing the organizational structure of reading to provide for individual differences.

In science education, Watson and Cooley¹⁶ have indicated that research is needed to better understand the learning process, the learner, and the teacher. More specifically, they argue that research is needed to understand "the conceptual processes by which young children form categories" and to develop new methods of measuring achievement in science.

Suggestions for Research in Other Areas of Education

It is not possible here to examine all of the suggestions for the selection of research topics. All we have attempted to do is to sample some of the suggestions made by specialists in the hope that the student will gain some insight into the type of research which is most needed in education. Students interested in suggestions related to administrative research, for example, should see Roesch,¹⁷ Stein,¹⁸ and Cocking.¹⁹ In research related to teacher education, the AACTE Study Series, *Needed Research in Teacher Education*,²⁰ should prove to be of value to interested students. In elementary education, Hendrickson,²¹ Smith,²² and LaBrant, Marcus, and Steinberg²³ provide students with possible projects. Students interested in secondary education will find the article by Angelino²⁴ of some interest.

¹⁵William S. Gray, "Needed Research in Reading," *Elementary English*, XXXIX, No. 2 (1952), 100-108.

¹⁶Fletcher G. Watson and William W. Cooley, "Needed Research in Science Education," in *Rethinking Science Education, 59th NSSE Yearbook*, Part I, (Chicago: University of Chicago Press, 1960), pp. 297-312.

¹⁷Winston Roesch, "Selected Areas for Research in School District Reorganization," *Phi Delta Kappan*, XL, No. 8 (1959), 328-330.

¹⁸Harry L. Stein, "Needs and Dimensions of Research," *Phi Delta Kappan*, XXXVII, No. 7 (1956) 316-327.

¹⁹Walter Cocking, "Need for School Plan Research," *School Executive*, LXXVI, No. 1 (1956), 7.

²⁰Joint Committee of the American Association of Colleges for Teacher Education and the American Educational Research Association, *Needed Research in Teacher Education*, (Oneonta, N. Y.: American Association of Colleges for Teacher Education, 1954).

²¹Gordon Hendrickson, "Some Needed Research in Elementary Education," *The Elementary School Journal*, LI, No. 3 (1950), 127-135.

²²Nila Banton Smith, "Areas of Research Interest in the Language Arts," *Elementary English*, XXXIX, No. 1 (1952), 31-34, 50.

²³Lou LaBrant, Fred Marcus, and Erwin R. Steinberg, "Needed Research in Language Expression," *Elementary English*, XXIX, No. 1 (1952), 35-38.

²⁴Henry R. Angelino, "Needed Research," *Review of Educational Research*, XXX, No. 1 (1960), 86-88.

DELIMITING AND SPECIFYING THE PROBLEM

So far in this chapter we have described a number of criteria for selecting a research problem and have indicated that there are various sources for locating published research ideas. Many students, however, have no idea of any research topic which embraces their own interests and abilities, and which at the same time has the potential of becoming a useful contribution to educational theory or practice. In this section we will suggest more specific ways for the student to find a subject worthy of investigation.

Traditionally, empirical studies of education have been divided into a number of categories, such as educational psychology, elementary and secondary curriculum and methods, and educational administration. Each division, although not entirely independent, is sufficiently homogeneous in content and method so that the student will sooner or later have to commit himself primarily to one of them. Each of the major fields of interest can be further analyzed into various specializations such as the following:

<i>Educational Psychology</i>	<i>Elementary and Secondary Education</i>	<i>Educational Administration</i>
learning theory	objectives of the school	federal, state, and local aspects of administration
measurement	organization of classes	school boards
exceptional children	curriculum planning	school law
counseling and guidance	role of the teacher	finance
growth and development	relation of school to community	construction and maintenance
personality and character	specific areas of the curriculum	business management
	teaching methods	transportation facilities
		organization of schools
		pupil and teacher administration

Let us suppose that the student has selected educational psychology as his major field of interest and that measurement is his specialty. Here again, a further analysis can be made:

Major field of interest

Specialty

Possible areas of concentration

Educational Psychology

Measurement

1. Construction of teacher-made examinations
2. Statistical analyses of tests and test results

3. Evaluation of tests (reliability, validity, etc.)
4. Measurement of intelligence, aptitude, achievement, interest, or personality

Within educational measurement, for example, the student might wish to work on a problem concerning the construction of teacher-made examinations, the statistical analyses of tests and test results, or the measurement of personality. But what are the major problems which need investigating in, let us say, the construction of teacher-made examinations or the measurement of intelligence?

As the student reads the literature concerning intelligence testing, he should keep in mind that problems may be selected to test various theoretical concepts, to clarify methodology, or to resolve practical field problems. However, he would do well to begin his analysis with an examination of the various theoretical concepts which have been expressed concerning his area of concentration.²⁵ By doing so he can develop an overview of the field.

Suppose that a student is interested in intelligence testing and wishes to examine various theories of intelligence in the hope that a possible research topic will be forthcoming. Where does he begin? One useful source is that of textbooks in educational measurement. Usually texts are written in understandable language and help to identify the most important variables and concepts. In addition, most texts provide the reader with a bibliography, which is often helpful in acquiring further information.

Another excellent source for beginning the study of any educational topic is the *Encyclopedia of Educational Research*.²⁶ By using the index located in the center of the *Encyclopedia*, the student can find various articles dealing with his possible area of concentration. In general, in each of the articles in the *Encyclopedia* the author tries to provide the reader with an overview of the major problems and methods which have been used in the past. Thus, in the article titled "Aptitudes," the author defines the term, indicates some of the limitations of measures of general intelligence, considers intelligence in terms of differential abilities, and describes tests of special aptitudes such as "auding" [listening and attending], music, and art. At the end of the article the student will find a most helpful bibliography.

One difficulty with the *Encyclopedia of Educational Research* is that new editions are published approximately every ten years. Hence, much of the content cannot be kept up-to-date. However, because the *Encyclopedia* provides a general introduction to many topics, it is still recommended as

²⁵Because so few theoretical positions have been developed in education, the student may find it difficult to begin his review of the literature by referring to a body of theory. Nonetheless, the student still has the responsibility of showing the relationship between his proposed investigation and related studies which have already been completed.

²⁶Chester W. Harris, ed., *Encyclopedia of Educational Research* 3rd ed., (New York: The Macmillan Company, Publishers, 1960).

one of the first sources researchers should examine as they attempt to build up a review of the literature.

Another excellent source that should be consulted is the *Review of Educational Research*, which is published by the same organization that publishes the *Encyclopedia of Educational Research* (the American Educational Research Association). The *Review* attempts to survey the major research findings in the more active research areas of education every three years. Less active areas of research are covered every six years. It thus provides much of the same type of information provided by the *Encyclopedia of Educational Research*, but does not attempt to be cumulative. The *Review* covered educational and psychological testing in February, 1959, in February, 1962, and again in February, 1965.

Suppose, however, that the student wants even more up-to-date information than the last issue of the *Review* can provide. At this point he should consult the *Education Index*. Published ten times annually, the *Index* provides the most recently published articles in journals, books, yearbooks, and bulletins concerned with education. After July 1, 1961, personal author entries were eliminated in the *Index*, so the student now has to look up articles by topic and subject rather than by author.

It is not possible to indicate here all of the journals which may be of use to students in education who are trying to select research problems. Most universities have research librarians who specialize in helping students find the sources they need, and the student should take advantage of their services. In addition, Alexander and Burke²⁷ have published an excellent guide for the location of educational information and data. It provides a simple but rather complete analysis of the types of practical problems which researchers are bound to come across and supplies useful methods of gathering the needed information.

An example of how a research topic may be selected will be useful at this point. Suppose that the student has read the various references in the *Encyclopedia of Educational Research* concerning the meaning and measurement of intelligence. He understands various theoretical viewpoints concerning the concept. In addition, he may become familiar with much of the terminology needed to comprehend studies involving intelligence, such as *mental age*, *individual tests*, and *PA norms*. Then, by using the appropriate edition of the *Review of Educational Research*, in this case "Educational and Psychological Testing," the student can delimit his area of investigation. For example, the 1962 *Review* made the following suggestions for needed research:

Studies assessing the effect of long-term changes in individuals—the stability problem—have been less common [than studies assessing short-term fluctuations in test performance]. . . .

²⁷Carter Alexander and Arvid J. Burke, *How to Locate Educational Information and Data* (4th Ed.) (New York: Bureau of Publications, Teachers College, Columbia University, 1958).

Follow-up studies with the Wechsler tests have been rare. Duncan and Barrett (1961) reported a correlation of .82 over a 10-year period for a group of superior adults, but they used only 28 cases.

Studies relating test performance to various aspects of learning continued to be plagued with the problem of adequate measures of learning. This research area deserves the increasing attention it has been getting.

It is not known whether children whose performance scores are considerably higher than their verbal scores profit more from remedial or therapeutic experience than children with comparable verbal scores whose performance scores are not deviant.

The *Davis-Ecls Games*, constructed especially to provide a "fairer" measure for low socio-economic groups, was found to correlate negligibly with social status in a study by Noll (1960); but in another study by Knief and Stroud (1959), which was consistent with most previous findings, the *Davis-Ecls Games* discriminated according to social class about as much as other tests.²⁸

Each of these points could very well provide the student with a research topic in educational measurement. The scope of each has been delimited by making *specific* recommendations for possible research studies.

SUMMARY

1. Four criteria for the selection of a research project were described:
 - a. *Interest*. In its extreme form the criterion of interest assumes that all activities develop from the researcher's personal preferences. This point of view denies the public nature of science, which demands that research activities be the property of all concerned individuals. The researcher has an obligation to select problems which not only concern himself but which also have some *probability* of concerning other members of the scientific community.
 - b. *Economics*. The economic criterion for the selection of a research problem reminds the investigator that some proposed projects may not be feasible because of the expenses involved. It by no means states that expense should be the only criterion for assessing the value of a study.
 - c. *Ability and training*. Students should consider their own training before they commit themselves to a particular investigation. Both assets and liabilities should be examined as honestly as possible. The amount of time needed to gain competency in a given area should also be considered.
 - d. *Uniqueness*. Although no investigation can be entirely novel or unique, students should attempt to select projects which have the potential of adding new information to education. The criterion of uniqueness does not imply that studies cannot be redone to increase the extent to which their findings can be generalized.
2. The relationship between theory and the selection of a research problem was discussed. Empirical investigations can contribute to a better understand-

²⁸Norman E. Wallen, "Development and Application of Tests of General Mental Ability," *Review of Educational Research*, XXXII, No. 1 (1962), 15-24.

ing of a theory by clarifying, delimiting, and defining it; in addition, the consequences implied by a theory can be empirically investigated to help validate the theory.

3. Three types of research topics were described which could be useful in correcting methodological problems:

- a. Research topics may be selected to clarify apparently contradictory findings by other researchers.
- b. Many commonly accepted research findings are based upon faulty methodology. Students may find possible topics for research by redesigning such studies.
- c. Students may find valuable sources for possible research by examining the statistical analyses performed by other investigators and re-analyzing poorly treated data.

4. Research that is designed to investigate a narrow problem which is not related to theory has limited value, not only for the profession as a whole but also for the investigator himself. Such studies may yield knowledge which is of immediate consequence, but unless the study can be generalized to other persons, places, or events, its conclusions must of necessity be overly restricted.

5. Students may find published lists of possible research topics to be of value to them in choosing adequate problems for investigation. Examples of published suggestions were provided for research on measurement, gifted children, and various aspects of the curriculum.

6. The suggestion was made that the student attempting to select a research problem should first decide upon his major field of interest (e.g., educational psychology, elementary or secondary education, administration). Within this major field of interest, he can next select a "specialty," such as counseling and guidance, teaching of reading, or school finance. The specialty can then be further analyzed into possible areas of concentration.

A specific research topic can be selected, refined, and delimited to reasonable size and scope by examining the literature pertaining to a given "area of concentration." It was suggested that the use of the *Encyclopedia of Educational Research* would provide the student with a broad overview of a particular field of investigation and would also provide a useful beginning bibliography. The *Review of Educational Research* and the *Education Index* were also suggested as very useful sources for gathering information.

PRACTICE EXERCISES

1. Select any single issue from one of the following sources:

Journal of Educational Psychology

Educational Leadership

School Review

The Elementary School Journal

Journal of Experimental Education

- a. Examine carefully each empirical article in the issue you have selected and indicate how each can be justified. Which of the articles you have reviewed are most likely to make a valuable contribution to education? Why? Justify your position.
 - b. What criteria did the authors of the articles you have selected use in determining their topics to investigate?
2. Examine one issue of each of the following journals:
- Review of Educational Research*
Journal of Experimental Education
Educational and Psychological Measurement
- What types of research (analytic, descriptive, or experimental) does each journal contain?
3. Examine two or three journals in your own field of specialization. What are the current problems being investigated in your area? What theoretical positions are you aware of in your specialty? What types of applied problems are being studied in your field?
4. For each of the following fictitious titles of proposed studies, indicate whether or not you would be willing to have a master's degree candidate submit the title if you were chairman of his thesis committee. Justify your decisions. Indicate what additional information you might want. What improvements, if any, would you suggest?
- a. A Description of the Guidance Program at Elmville High School
 - b. Attitudes of High School Students Toward Elmville High School
 - c. A Follow-up of Students Graduating from Elmville High School
 - d. The Role of the Principal in Elmville High School
 - e. The Role of the Coordinator of Extracurricular Activities in California
 - f. A Technique for Using the Chalk Board More Effectively
 - g. The School Lunch Line: A Proposed Method for Increasing Its Efficiency
 - h. The Attitudes of School Board Members to Ability Grouping
 - i. The Effect of Monetary Rewards in Increasing Motivation

SELECTED RELATED READINGS

1. Bachrach, Arthur J., *Psychological Research: An Introduction*. New York: Random House, Inc., 1962, 113 pages. Chapter 1 (pp. 3-25) is a delightfully written account of how accidental discovery rather than active search has led to the most useful research problems. The student should keep in mind, however, that the scientists were actively engaged in searching for the answer to a specific problem when they accidentally came upon the solution to an even more important one.
2. Findley, Warren G., "The Impact of Applied Problems on Educational Research," in *First Annual Phi Delta Kappa Symposium on Educational Research*. Bloomington, Ill.: Phi Delta Kappa, Inc., 1960, pp. 43-53. A thoughtful consideration of how applied or "practical" problems

affect educational research. The author defends the notion that research findings must be applied to and evaluated in local classrooms before they can be generally accepted. In defending his point of view, Findley describes some of the requirements for performing practical research (see especially pp. 49–53). The student should also examine pp. 96–112 for a critical discussion of Findley's viewpoint.

3. Rummel, J. Francis, *An Introduction to Research Procedures in Education*. New York: Harper & Row, Publishers, 1958, 413 pages. See especially pp. 20–39. Rummel discriminates clearly between subjects suitable for term papers, field studies, theses, and dissertations. In addition, the criteria of interest, significance, ability, and availability of data are discussed.
4. Van Dalen, Deobold B., *Understanding Educational Research*, 2nd Ed., New York: McGraw-Hill Book Company, 1966, 525 pages. See especially pp. 119–144. A general introduction to the analysis of a problem. Especially valuable are the author's examples of how problems may be analyzed and evaluated.

CHAPTER FOUR

Reviewing the Literature

Selecting a research problem requires that students have some familiarity with what previous investigations have found. Many thesis chairmen have had the experience of graduate students selecting research problems which reflect a serious lack of understanding of what other investigators have done. This lack of understanding may exhibit itself in a number of ways. The student may be unable to justify the need for a proposed investigation. He may select a particular hypothesis to test which has already been confirmed or disconfirmed. Or he may use techniques which previous research has invalidated for his particular purpose. In any of these instances, a *review of the literature* might have pointed out the need for specific studies and suggested those techniques which might be most useful for some given problem.

When the student has an idea for a research project, the review of the literature helps him select a problem, delimit its size and scope, show its relationship to previously completed research, and examine tools and methods which may be of value in designing his study. With this information, the student is in a position to evaluate for himself the efficacy of his plans and the importance of his investigation. Without it, he may find himself wasting much time and effort.

Understanding the Relationship of the Proposed Topic to Previously Completed Studies

One major advantage of a review of the literature is that it organizes a proposed field of investigation and indicates the ways in which the proposed study is related to previous research. A well-organized review shows the relationship of various studies to each other and thus forms the initial groundwork for the development of theory. In addition, by describing the relationships that his proposed investigation has to completed studies, the educational researcher has the opportunity to demonstrate the possible contributions his investigation can make to a body of accumulated knowledge. Without a review of the pertinent literature, it would be quite difficult to

assess the potentiality of a given research proposal. In addition, of course, many problems have roots in diverse fields both within and outside of education; the investigator has the responsibility of indicating the connections—tenuous though they may be—between his studies and previously completed ones.

Analysis of Useful Methodology

The review of the literature helps the student to find research techniques that are useful in his own particular area of investigation. Almost all published studies describe methods and techniques used to gather information. By examining these studies in detail the student can provide himself with knowledge of how other investigators attacked problems similar to those which he may face.

Suppose, for example, that an investigator is interested in studying how prejudice develops among children. At what ages do children begin making differential responses favoring one racial group over another? Does prejudice develop in the same manner for members of minority groups as it does for members who are in the majority? The most apparent method of answering these questions might be to follow a group of Negro and Caucasian children throughout their childhood years. By observing their responses to members of other races, the researcher would have some idea when prejudice was beginning. The difficulty with that approach, of course, is the amount of time it would take.

In reviewing the literature on the development of prejudice, the student is likely to come across a study by the Clarks,¹ which employed dolls that were identical except for color. By asking Negro children questions about the dolls, the experimenters were able to gather information concerning preferences, knowledge of racial differences, and extent to which the children identified with members of their own racial group. This particular technique might also prove to be of value in other studies involving the genesis of prejudice, ego involvement, authoritarianism, or stereotypes.

WRITING THE REVIEW OF THE LITERATURE

Students often consider the writing of a review of the literature as an unnecessary requirement designed by their professors to make them prove that the proposed research is an original contribution. Having this belief, the student begins his review with something less than the zeal which the task demands.

¹Kenneth B. Clark and Mamie P. Clark, "Racial Identification and Preference in Negro Children," in *Readings in Social Psychology*, ed. Eleanor E. Maccoby, Theodore M. Newcomb, and Eugene L. Hartley (3rd ed.), (New York: Holt, Rinehart & Winston, Inc., 1958), pp. 602-611.

The review of the literature is not designed to prove the originality of a research design or topic. A review which attempts to prove that no other person has considered a particular study before is wasteful and unnecessary. Far from showing uniqueness, the review of the literature primarily should show how the proposed study is related to other investigations. In addition, it should, of course, show the contribution to a particular field that the proposed research will make.

A second misconception concerning the review of the literature is that it forces the investigator to examine every study which is at all related to his proposed research. The student may thus consider the review as an attempt by his professors to compel him to be scholarly and complete. There are two major arguments which may be raised against an exhaustive review of the literature. The first is that it would be impossible. Literally hundreds of studies exist on any given topic. To be forced to read each would require a lifetime of devotion just to write the review of the literature. Considering the interrelatedness of investigations, an exhaustive search on any one topic would no doubt turn up far more studies than one would possibly need to justify a proposed research and to show its relationship to other investigations. This point is related to the second argument against an exhaustive search. Even if it were possible to examine every study related to a given topic, it would still be unnecessary to do so. The purpose of the review is to show how the proposed research is related to previous investigations and how it can make a unique contribution. These objectives do not require that all possible sources be included in the review. They do mean that *the review must be selective rather than exhaustive and organized rather than a list of published investigations, and that it should show the relationship of the proposed study to other studies.*

Selection of Studies to Be Included in the Review of the Literature

Making the review of the literature selective rather than exhaustive requires that the student be discriminating in his choice of articles and books to be included in the review. The suggestions which follow are designed to help the student select the most relevant sources of information.

- a. Many more references can be included in the review of the literature in a thesis or dissertation than can be reported if the research is to be published in a journal. Journal space is usually quite limited, making it necessary to include only the most relevant information. In addition, journals are usually designed to be read by specialists, who presumably are familiar with previous research studies; therefore, the review can be curtailed. In master's theses and doctoral dissertations, however, the paper is usually addressed to the intelligent layman who has had limited training in the particular field of investigation.
- b. Of two similar and equally valuable articles on the same subject, select the more recent for detailed analysis. The older article may be referred to

by stating: "Similar results were found by Jones (1927)." In this way, the review is kept as up-to-date as possible, but credit is still given to earlier relevant works.

- c. Some indication should be made why each reference in the review of the literature was included. The review may include some references to articles because of their historical interest. Thus, the reviewer might state: "Although Pavlov (10) appears to have been the first investigator to employ the differential effect of partial versus continuous reinforcement, it was Skinner (12, 46) who recognized the theoretical and practical importance of Pavlov's discoveries."
- d. Some articles might be included in the review because the writer wishes to discriminate between various theoretical positions or between findings that are ostensibly contradictory. Where this is the case, the writer can describe the major points of view for each position, and then simply provide a reference to the advocates of each. Thus the review of the literature might state, "Whereas Skinner (1958) has advocated the required use of many small steps in the development of material suitable for programmed instruction, Crowder (1960) has suggested that some students may skip materials which are simple for them. In partial support of Skinner's contention, Coulson and Silberman (1960) found that the use of small steps led to significantly higher scores on examinations than did the use of larger steps."
- e. Where there is some disagreement in opinion or methodology in the published literature, the student can select articles for inclusion in his review which seem to be representative of various points of view or which have used different techniques.

Organization of the Review of the Literature

The review of the literature should never be a simple listing of various studies which are vaguely related to the proposed topic for investigation. Each review should be organized; this organization should be made explicit by dividing the review into major and minor headings.

In organizing the review, one must adhere to the logical principles of division. These principles require that the student develop a system in much the same way that he would develop an outline, by beginning with major headings (usually given Roman numerals I, II, III, IV, etc.) and subdividing each major heading into subparts (A, B, C, D). If further subdivisions are necessary, they may be given lower-case Arabic numerals (1, 2, 3, 4), and these may be further divided by using lower-case letters (a, b, c, d). For example, the concept of measurement could be partially outlined as follows:

MEASUREMENT

I. Psychological Measurement

A. Types

1. Intelligence

- a. individually administered tests
- b. group tests

- 2. Aptitude
- 3. Personality
 - a. projective techniques
 - (1) Thematic Apperception Test
 - (2) Rorschach
 - b. pencil-and-paper tests
- B. History
- II. Physical Measurement
 - A. Weight
 - B. Height
 - C. Distance

The *principle of exhaustiveness* requires that the outline be complete. For example, points I and II should be the only types of measurement possible. Similarly, IA and IB should contain all possible knowledge concerning psychological measurement, and points 1, 2, and 3 would have to include all types of tests. By not including attitude scales under section IA, the fallacy of *incomplete division* is committed.

The *principle of exclusiveness* requires that categories within headings do not overlap each other. Thus, dividing types of schools into preschools (ages 5 and 6), elementary schools (grades 1-6), junior high schools (grades 7-9), and high schools (grades 10-12) would violate this principle, because some children who are 6 years of age might be in the first grade. These children would, therefore, have more than one classification which refers specifically to them.

The *principle of single classification* requires that all items included at a given level be selected on the same basis. For example, in the outline given above, physical measurement was subdivided according to a single type of classification, object of measurement. If subtopic D had been "Pounds," then the fallacy of *cross division* would have been committed. The use of "Pounds" would change the topics under physical Measurement to include two bases for selection: object of measurement and unit of measurement. Similarly, the classification of schools into levels is perfectly legitimate as long as a fourth level such as "vocational education" is not added.

The reader of the review of the literature deserves to be told the basis upon which it is being organized. This is most simply accomplished by providing the reader with an introductory paragraph or two describing how the review is organized. For example, in one study² which compared physics students of various levels of abilities who were taught by traditional methods with students taught by the use of a television film, the dissertation writer began his review of the literature in the following way:

²Leonard James Garside, "A Comparison of the Effectiveness of Two Methods of Instruction in High School Physics as Measured by Levels of Achievement of Students of High and Low Intelligence," (Doctoral Dissertation, University of Wisconsin, 1959).